

IBM InfoSphere DataStage and QualityStage
Version 11 Release 3

Guide to Accessing Files



IBM InfoSphere DataStage and QualityStage
Version 11 Release 3

Guide to Accessing Files



Note

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 27.

Contents

File connector	1
Designing jobs that use the File connector	1
Defining a job that includes the File connector	1
Defining a connection	2
Reading data from a file	4
Writing data to a file.	5
Reference	8
File encoding	8
Runtime column propagation.	9
File name format	9
Limitations	12
Appendix A. Product accessibility	15
Appendix B. Reading command-line syntax	17
Appendix C. How to read syntax diagrams.	19
Appendix D. Contacting IBM	21
Appendix E. Accessing the product documentation	23
Appendix F. Providing feedback on the product documentation	25
Notices and trademarks	27
Index	33

File connector

Use the File connector to read files from and write files to a local file system on the engine tier or a Hadoop Distributed File System (HDFS) by using the WebHDFS API or the HttpFS API.

If you connect to HDFS, you might want to use the File connector instead of the Big Data File stage for the following reasons:

- The Big Data File stage supports only InfoSphere® BigInsights™, but the File connector supports HDFS by using the WebHDFS API and the HttpFS API.
- The File connector is independent of the HDFS version.
- The File connector supports Kerberos authentication.

Designing jobs that use the File connector

You can use the File connector to develop jobs that read data from and write data to files.

Procedure

1. Define a job that includes the File connector.
2. Define a connection to a file system.
3. To set up the File Connector stage to read data from a file, complete the following steps:
 - a. Set up column definitions.
 - b. Configure the File connector as a source.
 - c. Optional: Create a reject link to manage rejected data.
 - d. Optional: Read partitioned data.
4. To set up the File Connector stage to write data to a file, complete the following steps:
 - a. Set up column definitions.
 - b. Configure the File connector as a target.
 - c. Optional: Write partitioned data.

Defining a job that includes the File connector

To read data from or write data to a file, you can create a job that includes the File connector. Then, you add any additional stages that are required and create the necessary links.

Procedure

1. In the InfoSphere DataStage® and QualityStage® Designer client, select **File > New** from the menu.
2. In the **New** window, select the **Parallel Job** icon, and then click **OK**.
3. Add the File connector to the job:
 - a. In the palette, select the **File** category.
 - b. Drag the File Connector stage to the canvas.
 - c. Optional: Rename the File Connector stage. Choose a name that indicates the role of the stage in the job.

4. Create the necessary links and add additional stages for the job:
 - For a job that reads a file, create the next stage in the job, and then create an output link from the File connector to the next stage.
 - For a job that writes data to a file, create an input link from the previous stage in the job to the File connector.
5. Save the job.

Defining a connection

You can access the local file system on the engine tier or access a Hadoop Distributed File System (HDFS) by using the WebHDFS API or the HttpFS API.

About this task

To access HDFS, you must define a connection that specifies the server name, server port number, user name, and password. Alternatively, you can also specify the connection URL instead of the host name and port.

You can use Kerberos authentication to connect to HDFS by using the WebHDFS API or the HttpFS API. If you use Kerberos authentication, you must specify the Kerberos principal in the **User name** property and the password in the **Password** property. Optionally, you can also specify the realm of the principal in the user name property, for example, `principal@realm`. Alternatively, you can use the Kerberos keytab file instead of the password.

When Kerberos authentication is used, the File connector uses the `krb5.conf` file. Specify the Kerberos Key Distribution Center (KDC) host name and the default realm in the `krb5.conf` file. For more information about the `krb5.conf` file, see the Kerberos documentation.

Procedure

Configure a connection to a local file system or HDFS by using the WebHDFS API or the HttpFS API.

Option	Procedure
Access the local file system	<ol style="list-style-type: none"> 1. On the Properties page, in the Connection section, set the File system property to Local. 2. Click OK.

Option	Procedure
Access a file system on HDFS by using the WebHDFS API or the HttpFS API	<ol style="list-style-type: none"> 1. On the Properties page, in the Connection section, set the File system property to WebHDFS or HttpFS. 2. Specify the host name, user name, and the password for the user. 3. Specify a port number. If a port number is not specified, the connector uses one of the following port numbers: <ul style="list-style-type: none"> • If the Use SSL (HTTPS) property is set to No, the connector uses 50070 (WebHDFS) or 14000 (HttpFS). • If the Use SSL (HTTPS) property is set to Yes, the connector uses the port number 50470 (WebHDFS) or 14443 (HttpFS). 4. Click OK.

Configuring the truststore

To use SSL encryption when you use the WebHDFS API or HttpFS API to communicate with the Hadoop Distributed File System (HDFS), you might need to import the server public certificate into your truststore and specify values for truststore parameters.

About this task

You must configure the truststore in the following scenarios:

- The server instance that you connect to uses a self-signed certificate.
- The server instance that you connect to uses a certificate that is issued by an authority that is not contained in the default Java truststore.
- You do not want to use the default Java truststore. The File connector can use the default Java truststore, a custom truststore, or both.

Procedure

1. Log in as the administrator user to the computer where the InfoSphere Information Server engine tier host is installed.
2. If a truststore does not exist, use the keytool Java™ utility to create a truststore by issuing the following command:

```
keytool -genkey -alias alias -keystore truststore.jks -storepass password
```

Where *truststore.jks* is the name of the truststore to create and *alias* is a keystore entry where certificates are stored. The value specified for *alias* must be unique. The default location for the keytool Java utility is *install_directory/jdk/bin/keytool*.

3. Move the server certificate (.pem) file to the computer where the InfoSphere Information Server engine tier is installed.
4. Use the keytool Java utility to import the server certificate into the File connector truststore by issuing the following command:

```
keytool -import -trustcacerts -alias alias -file pem_file -keystore truststore.jks -storepass password
```

5. To generate an encrypted version of the password, use the *IS_installation_directory/ASBNode/bin/encrypt.sh* file or *encrypt.bat* file.

6. Create a properties file in plain text format, for example, `properties.txt`, and add the following line:

```
password=password
```

Where *password* is an encrypted version of the password that was specified as the value for the `-storepass` parameter for the `keytool import` command.

7. In the `IS_installation_directory/Server/DSEngine/dsenv` directory or at the InfoSphere DataStage project level, set the following environment variables:

```
DS_TRUSTSTORE_LOCATION=path_to_the_truststore.jks_file
DS_TRUSTSTORE_PROPERTIES=path_to_the_properties.txt_file
```

Reading data from a file

You can configure the File connector to read data from a file.

Before you begin

- Define a job that contains the File Connector stage.
- Define a connection to a file system.

Setting up column definitions on a link

Column definitions, which you set on a link, specify the format of the data records that the connector reads from a data source or writes to a data source.

Procedure

1. From the job design canvas, double-click the connector icon.
2. Use one of the following methods to set up the column definitions:
 - Drag a table definition from the repository view to the link on the job canvas. Then, use the arrow buttons to move the columns between the **Available columns** and **Selected columns** lists.
 - On the **Columns** page, click **Load** and select a table definition from the metadata repository. Then, to choose which columns from the table definition apply to the link, move the columns from the **Available columns** list to the **Selected columns** list.
3. Configure the properties for the columns:
 - a. Right-click within the columns grid, and select **Properties** from the menu.
 - b. Select the properties to display, specify the order in which to display them, and then click **OK**.
4. Optional: Modify the column definitions. You can change the column names, data types, and other attributes. In addition, you can add, insert, or remove columns.
5. Optional: Save the new table definition in the metadata repository:
 - a. On the **Columns** page, click **Save**, and then click **OK** to display the repository view.
 - b. Navigate to an existing folder, or create a new folder in which to save the table definition.
 - c. Select the folder, and then click **Save**.

Configuring the File connector as a source

To configure the File connector to read file data from a local file system on the engine tier or a Hadoop Distributed File System (HDFS), you must configure properties for the file that you want to read.

Procedure

1. From the job design canvas, double-click the File Connector stage.
2. In the **File name** property, specify the name of the file to read.
3. Optional: In the **Exclude files** property, specify the name of a file that contains a comma-separated list of prefixes that the files that you want to exclude must have in their file path.
4. In the **File format** property, specify the file formatting options.
5. Click **OK**, and then save the job.

Reading partitioned data

In a job that uses multiple nodes, each node that is specified for the stage reads a distinct subset of data from the source.

Before you begin

- Configure the stage on the output link to run in parallel.
- Define two or more processing nodes to run the job on.

About this task

The File connector can read one or more files sequentially or in parallel. To read more than one file, specify wildcards in the value for the **File name** property.

When the job runs in sequential mode, the connector reads all rows from all input files by using a single processing node. When the job runs in parallel mode, each input file whose name matches the specified wildcard options, is read by multiple processing nodes. Each processing node reads a subset of rows from every input file.

If you want each node to read one or more different files, the specified file names must contain a unique number that corresponds to a node number. Node number is the value that is specified for the **File name** property. For example, if you define two nodes for the job and specify `MyFile_[[node-number]].txt` for the **File name** property, node 0 reads the `MyFile_0.txt` file and node 1 reads the `MyFile_1.txt` file.

Use the `[[node-number]]` option and wildcards in the **File name** property to combine the two methods for each processing node to read all rows from a distinct set of matching input files.

Procedure

1. On the job design canvas, double-click the File Connector stage, and then click the **Stage** tab.
2. On the Advanced page, set **Execution mode** to **Parallel**.
3. Specify a value for the **File name** property based on the number and names of the files to read.
4. Click **OK**, and then save the job.

Writing data to a file

You can configure the File connector to write data to a single file, or multiple files, or delete files.

Before you begin

- Define a job that contains the File Connector stage.
- Define a connection to a file system.

Setting up column definitions on a link

Column definitions, which you set on a link, specify the format of the data records that the connector reads from a data source or writes to a data source.

Procedure

1. From the job design canvas, double-click the connector icon.
2. Use one of the following methods to set up the column definitions:
 - Drag a table definition from the repository view to the link on the job canvas. Then, use the arrow buttons to move the columns between the **Available columns** and **Selected columns** lists.
 - On the **Columns** page, click **Load** and select a table definition from the metadata repository. Then, to choose which columns from the table definition apply to the link, move the columns from the **Available columns** list to the **Selected columns** list.
3. Configure the properties for the columns:
 - a. Right-click within the columns grid, and select **Properties** from the menu.
 - b. Select the properties to display, specify the order in which to display them, and then click **OK**.
4. Optional: Modify the column definitions. You can change the column names, data types, and other attributes. In addition, you can add, insert, or remove columns.
5. Optional: Save the new table definition in the metadata repository:
 - a. On the **Columns** page, click **Save**, and then click **OK** to display the repository view.
 - b. Navigate to an existing folder, or create a new folder in which to save the table definition.
 - c. Select the folder, and then click **Save**.

Configuring the File connector as a target

You can configure the File connector to write rows to or delete rows from the file.

Procedure

1. On the job design canvas, double-click the File Connector stage.
2. Select the input link to edit.
3. Configure the connector to write rows to or delete rows from a file.

Table 1. Procedures for writing or deleting data from a file

Action	Procedure
Write single file	<ol style="list-style-type: none">1. Set the Write mode property to Write single file.2. Specify the name of the file to write.

Table 1. Procedures for writing or deleting data from a file (continued)

Action	Procedure
Write multiple files	<ol style="list-style-type: none"> 1. Set the Write mode property to Write multiple files. 2. Specify the base name of files to write. The connector customizes the name for each file it writes by using the specified base name as a template.
Delete	<ol style="list-style-type: none"> 1. Set the Write mode property to Delete. 2. Specify a text column on the input link. The connector deletes the file that is named in each row of the column.

4. In the **File format** property, specify the file formatting options.
5. Click **OK**, and then save the job.

Writing partitioned data

In a job that uses multiple nodes, you can configure the File connector to run in parallel and generate multiple files per processing node.

Before you begin

- Configure the stage on the input link to run in parallel.
- Define two or more processing nodes to run the job on.

About this task

When you configure the File connector to write in parallel, each node writes data to a separate file or files. By default, the connector generates file names in the following format:

value_in_file_name_property.key-value.part00000.000

value_in_file_name_property

The name that is specified in the **File name** property.

key-value

The specified key column value.

part00000

The partition number, which is a processing node number that starts with 0.

000

The file sequence number. Every time that a processing node creates a file, the file sequence number increments by one. This number is not reset for each key value. If a processing node creates five files, the sequence numbers in the file names range 0 - 4 regardless of the number of key value changes.

For example, if the data has six key values and the stage uses three processing nodes, the file names might be as follows:

```
myfile.key0.part00000.000
myfile.key0.part00000.001
myfile.key1.part00000.002
myfile.key1.part00000.003
myfile.key2.part00001.000
myfile.key2.part00001.001
myfile.key3.part00001.002
myfile.key3.part00001.003
```

```
myfile.key4.part00002.000
myfile.key4.part00002.001
myfile.key5.part00002.002
myfile.key5.part00002.003
```

If a key column value is not specified, the connector generates files for three nodes in the following file name format:

```
myfile.part00000
myfile.part00001
myfile.part00002
```

Procedure

1. On the job design canvas, double-click the File Connector stage, and then click the **Stage** tab.
2. On the Advanced page, set **Execution mode** to **Parallel**.
3. On the Stage page, configure the write mode properties.
 - a. Set the **Write mode** property to **Write multiple files**.
 - b. Specify a value for the **File name** property.
 - c. Optional: Specify a value in megabytes for the **Maximum file size** property. When a file reaches the maximum file size that is specified, the connector creates a separate file.
 - d. Specify a value for the **Split file on key changes** property to indicate whether to create a new file when the value in the key column changes.
 - a. Specify the name of the key column in the **Key column** property.
4. On the Input page, select the Partitioning page to configure partitioning properties.
 - a. Set the **Partition type** property to **Hash** and use the same key column that is specified in the Properties page.
 - b. Select the **Perform sort** check box to enable sorting of rows and use the same key column name that is specified in the Properties page. Make sure that the case sensitivity of the column that is identified for sorting matches the value that is specified in the **Case sensitive** property on the Properties page.
5. Click **OK**, and then save the job.

Reference

To use the File connector successfully, you might need detailed information, such as information about patterned writes, place holders, and file formats.

File encoding

You can control the encoding of the files that the File connector reads or writes.

You can specify the file encoding in the following ways, which are listed in their order of precedence:

1. As a value for the **Encoding** property in the stage editor.
2. As a value for the charset attribute in a `.osh` schema file. You can use this method only if runtime column propagation is enabled and the connector uses metadata from a `.osh` schema file.
3. As a value for the **APT_IMPEXP_CHARSET** environment variable.

To specify the file encoding, use the **Encoding** property. You can also specify the character set encoding of files in for the charset attribute in the `.osh` schema file or

the **APT_IMPEXP_CHARSET** environment variable. If a value is specified in the **Encoding** property, the connector ignores the values in the charset attribute in the .osh schema file and the environment variable. If the property is not set, the connector uses the value in the .osh schema. If neither are set the connector uses the value in the environment variable.

You can use the job NLS map or the project NLS map to specify the character set for columns in your job of the Char, VarChar, and LongVarChar columns data types. When you use the File connector to write data, the value that is specified in the NLS map is used for columns of that the Char, VarChar, and LongVarChar data types. When you use the File connector to read data, connector encodes Char, VarChar, and LongVarChar data types in the specified encoding. Columns of the NChar, NVarChar, and LongNVarChar data types are always encoded in UTF-16.

To ensure that data is not corrupted, use the same encoding for columns of Char, VarChar, and LongVarChar data types in all stages in the job. For more information about NLS configuration for other stages, see the related documentation.

Runtime column propagation

Use runtime column propagation to have the connector automatically add missing columns to the link schema when the job runs.

Usage

Before you can enable runtime column propagation in a stage, runtime column propagation must be enabled for parallel jobs at the project level from the InfoSphere DataStage Administrator client. To enable runtime column propagation for the output link of the stage, select the **Runtime column propagation** check box on the Columns page.

When runtime column propagation is enabled, the connector uses the column definition that is specified in one of the following ways by using the **Record definition** property:

- As the first row or file header of the file that is read
- As a delimited string that is specified as the value for the **Definition source** property
- As a delimited string in the file that is specified as the value for the **Definition source** property
- As a .osh schema file that is specified as the value for the **Definition source** property

File name format

You can specify the file name as an absolute path or a relative path. You can also use wildcards and place holders in the file name.

For local file systems, the path is relative to the InfoSphere DataStage project directory. For HDFS file systems, the path is relative to the home directory of the user.

Pattern matching based on wildcard characters

In read mode, you can use wildcard characters in the **File name** property. The connector uses the glob syntax for pattern matching.

Wildcard characters

You can use any of the following wildcard characters:

- * Specify asterisk (*) to match zero or more characters. For example, specify *.txt to match all files with the .txt extension.
- ? Specify a question mark (?) to match one character.

[range]

Specify a range to match a single character in a set or a range of characters. For example, you can specify the range [a-f] to match [abcdef].

If multiple files match the pattern, the connector reads each file, one after another. When jobs run in parallel, each processing node reads a different part of the matching file or files.

File name format options

You can customize the format of the file name that the File connector reads or writes by using one or more placeholders, in the file name that you specify in the **File name** property. The placeholders are replaced with actual values when the job is run.

The following table shows the supported placeholders and the context in which they can be used .

Table 2. File name options for reading and writing partitioned data

Placeholder	Context	Description
[[node-number(spec)]]	Read, Write	The number of the processing node. If a job runs in parallel and the node number is not specified, the connector automatically appends the node number. You can optionally specify a specification. For more information, see the Specifications section.
[[node-count]]	Read, Write	The total number of processing nodes.
[[sequence-number(spec)]]	Write	The file sequence number. Every time that a processing node creates a file, it increments the sequence number by one. If this option is not specified as part of the value for the File name property, the connector adds the sequence number in the default position. You can optionally specify a specification. For more information, see the Specifications section.

Table 2. File name options for reading and writing partitioned data (continued)

Placeholder	Context	Description
[[key-value]]	Write	The key column value. If this option is not specified as part of the value for the File name property and the Use key value in file name property is set to Yes, the connector adds the key value in the default position.

Specifications

When you include node numbers or sequence numbers in the file name, you can specify how to format those numbers. The first node can be node 0, 1, or 0 or 1 with leading zeros. By default, the first node is node 0. For example,

- To start the node number from 1, specify 1.
- To start a node number 0 or 1 with three leading 0s, specify 0000 or 0001.

For example, if you want the file name that is read from or written by the first node (node 0) to be `MyFile.01.of.04`, and when there are four nodes, then specify this value for the **File name** property:

`MyFile. [[node-number(01)]] .of. [[node-count]]`

File name options for partitioned reads

The value that is specified for the **File name** property affects how the File connector reads partitioned data.

If the value that is specified for the **File name** property contains the `[[node-number]]` placeholder, each processing node reads all rows from a distinct input file. If wildcards are used, each processing node reads a subset of rows from multiple files that match the pattern. If both the placeholder and the wildcards are used, each processing node reads a distinct set of matching files.

The following tables show how you can configure the information that is included in the file name.

Table 3. Behavior of the File connector in the sequential mode with no placeholders in the file name

No wildcards	Wildcards
The File connector reads all rows from a single input file.	The File connector reads all rows from each matching input file.
File name: myfile.txt	File name: myfile*.txt
Pattern: myfile.txt	Pattern: myfile*.txt
Number of files read: One	Number of files read: Zero or more

Table 4. Behavior of the File connector in the sequential mode with placeholders in the file name

No wildcards	Wildcards
The File connector reads all rows from a single input file.	The File connector reads all rows from each matching input file.
File name: myfile.[[node-number]].of.[[node-count]]	File name: myfile*.[[node-number]].of.[[node-count]]
Pattern: myfile.1.of.1	Pattern: myfile*.1.of.1
Number of files read: One	Number of files read: Zero or more

Table 5. Behavior of the File connector in the parallel mode with no placeholders in the file name

No wildcards	Wildcards
Each processing node reads a subset of rows from a single input file.	Each processing node reads all input rows from a different set of input files.
File name: myfile.txt	File name: myfile*.txt
Pattern: Node 1: myfile.txt (Subset of rows) Node 2: myfile.txt (Subset of rows) Node 3: myfile.txt (Subset of rows) Node 4: myfile.txt (Subset of rows)	Pattern: Node 1: myfile*.txt (Subset of rows) Node 2: myfile*.txt (Subset of rows) Node 3: myfile*.txt (Subset of rows) Node 4: myfile*.txt (Subset of rows)
Number of files read: One per node	Number of files read: Zero or more per node

Table 6. Behavior of the File connector in the parallel mode with placeholders in the file name

No wildcards	Wildcards
Each processing node reads all rows from a different input file.	Each processing node reads all input rows from a different set of input files.
File name: myfile.[[node-number]].of.[[node-count]]	File name: myfile*.[[node-number]].of.[[node-count]]
Pattern: Node 1: myfile.1.of.4 (All rows) Node 2: myfile.2.of.4 (All rows) Node 3: myfile.3.of.4 (All rows) Node 4: myfile.4.of.4 (All rows)	Pattern: Node 1: myfile*.1.of.4 (All rows) Node 2: myfile*.2.of.4 (All rows) Node 3: myfile*.3.of.4 (All rows) Node 4: myfile*.4.of.4 (All rows)
Number of files read: One per node	Number of files read: Zero or more per node

Limitations

The File Connector has some limitations when creating Hive tables.

- You cannot create Hive tables when connected to an instance of InfoSphere BigInsights because of limitations of the JDBC driver.

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 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.

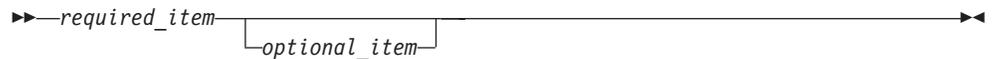
Appendix C. How to read syntax diagrams

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).



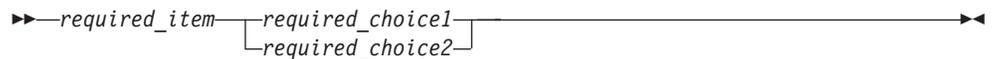
- Optional items appear below the main path.



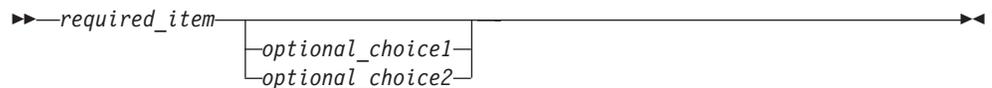
If an optional item appears above the main path, that item has no effect on the execution of the syntax element and is used only for readability.



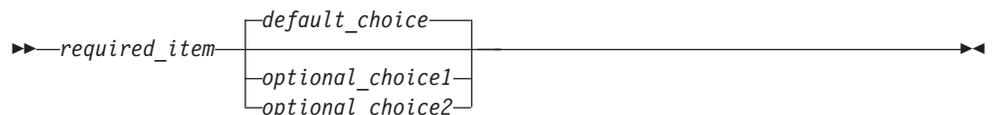
- If you can choose from two or more items, they appear vertically, in a stack. If you must choose one of the items, one item of the stack appears on the main path.



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.

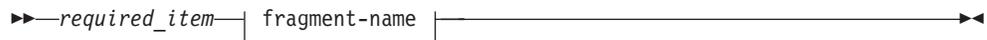


If the repeat arrow contains a comma, you must separate repeated items with a comma.



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 7. 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
```

Setting up a locally installed information center

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.

Set up a locally installed information center on the computer of your choice by completing one of the following actions:

- Use the installation instructions that come with the information center installation package
- Use the installation instructions in the following technote: <http://www-01.ibm.com/support/docview.wss?uid=swg27042237>

Changing help links to refer to locally installed documentation

After you install and start your locally installed information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product 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 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>.

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 8. 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 8. 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>.

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

Trademarks

IBM, the IBM logo, and 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

- command-line syntax
 - conventions 17
- commands
 - syntax 17
- Configuring
 - file connector as a target 6
- Connection to a file system
 - defining 2
- connector
 - column definitions 4, 6
- customer support
 - contacting 21

F

- file connector
 - examples
 - reading data 4
- File connector 1
 - configuration
 - configuring the File connector as a source 5
 - connections 3
 - examples
 - writing data 6
 - job definition 1
 - job design 1
 - overview 1
 - partitioned data
 - reads 5
 - reads
 - configuration 5
 - overview 4
 - parallel reads 5
 - partitioned reads 5
 - runtime column propagation
 - overview 9
 - writes
 - configuration 6
 - overview 6
 - parallel writes 7
- File connector File connector
 - partitioned data
 - writes 7
 - writes
 - partitioned writes 7
- File name format 9, 10
- File name options 10, 11
- File name property 10

L

- legal notices 27

P

- Patterns 10

- product accessibility
 - accessibility 15
- product documentation
 - accessing 23

R

- Reading partitioned data 10, 11

S

- security certificate
 - importing 3
- software services
 - contacting 21
- special characters
 - in command-line syntax 17
- support
 - customer 21
- syntax
 - command-line 17

T

- trademarks
 - list of 27

W

- web sites
 - non-IBM 19
- Writing partitioned data 10, 11



Printed in USA

SC27-6722-00

