Technical Bulletin

Part No. 00D-TB039

DataStage Row Splitter

This technical bulletin describes DataStage Row Splitter. This stage allows you to divide a single formatted input column into a number of separate output columns.

Printing History

First Edition for Release 1.0, December 2003

How to Order Technical Documents

To order copies of documents, contact your local Ascential subsidiary or distributor, or call our main office at (508) 366-3888.

Documentation Team: Mandy deBelin and Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 1.0 of the DataStage Row Merger stage for DataStage Release 7.0.1:

Row Splitter reads data one row at a time from an input link. It splits the data fields contained in a string into a number of columns. It then writes the columns to the output link. The stage can have a single input link and a single output link.

In normal operation of Row Splitter, each input string processed results in an output row of multiple columns. In some cases, however, a single input string may represent several rows of input data. In this case the stage can deconcatenate these into separate rows for output.

Note: Row Splitter is similar to the server sequential file stage. The difference is that, while the sequential file stage reads from a file, Row Splitter reads from a link.

This bulletin describes the user interface for Row Splitter.

Functionality

Row Splitter has the following functionality and benefits:

- Reads one row at a time, splits the data fields contained in a string into a number of columns, and then writes the columns to the output link.
- Supports generation of multiple output rows.
- Supports NLS (National Language Support). The stage writes what it reads without interpretation or conversion.

The following functionality is not supported:

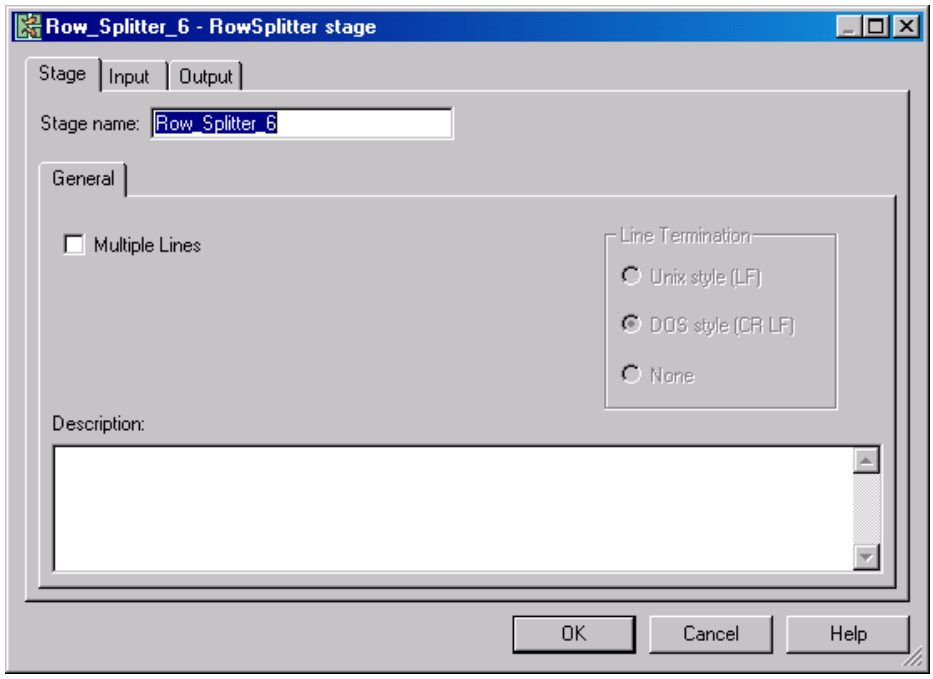
- Support for MetaStage.

Installing the Plug-in

For instructions and information supporting the installation, see *DataStage Plug-In Installation and Configuration Guide.*

Stage Page General Tab

The **General** tab of the **Stage** page gives access to the deconcatenation facilities of Row Splitter.



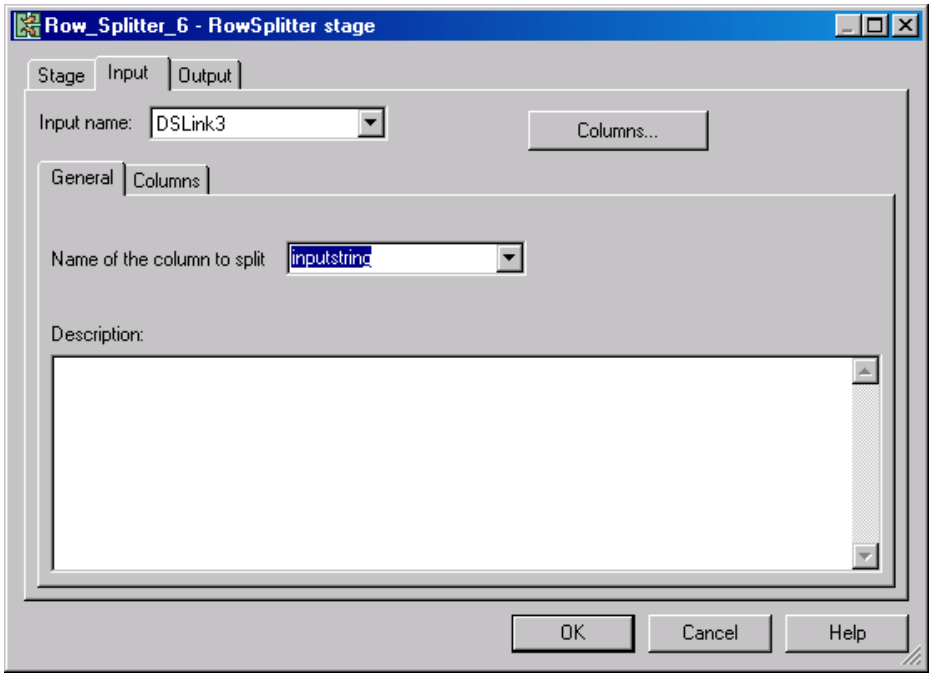
The **General** tab contains the following fields:

- **Multiple Lines**. This determines whether Row Splitter deconcatenates the input string into separate output rows, or whether it outputs each input string as a separate output row. Select **Multiple Lines** to have the rows deconcatenated. By default it is not selected.
- **Line Termination**. This setting is only available if you have chosen the **Multiple Lines** option to specify that the stage is deconcatenating input rows. It specifies the character(s) that are placed as a delimiter between the concatenated rows, so the stage knows where to split them. Choose between:
  - Unix Style (LF). The delimiter is a linefeed character.
  - DOS style (CR LF). The delimiter is a carriage return character and a linefeed character.
  - None. There is no delimiter.
- **Description**. Enter an optional description of the stage.

Inputs Page

The **Inputs** page contains various tabs that describe the rows of data being input to Row Splitter.

General Tab

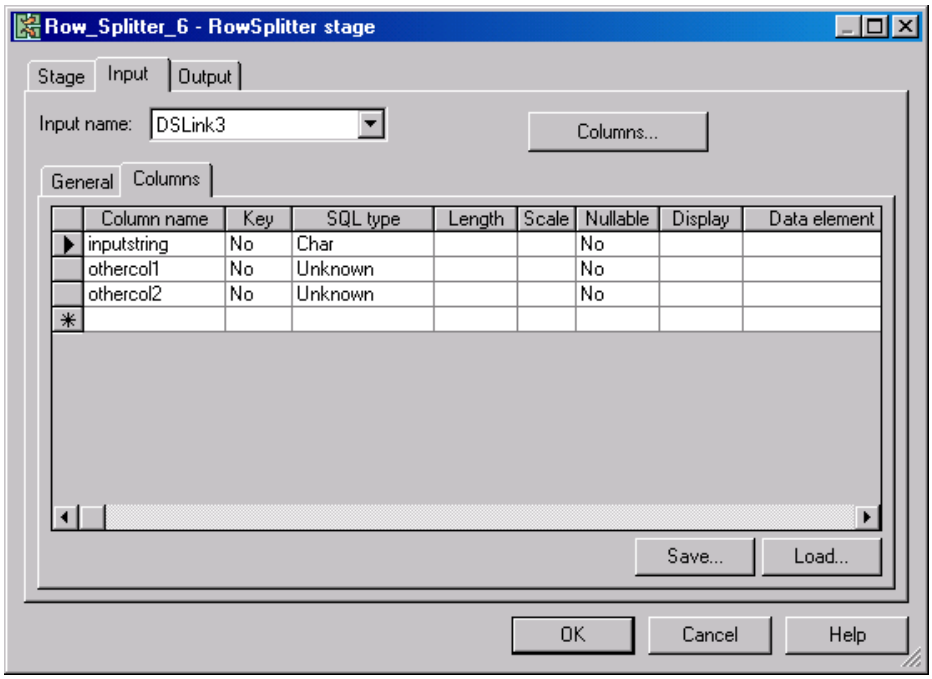


The **General** tab contains the following fields:

- **Name of the column to split**. A drop-down list contains a list of the defined input columns for this stage. Choose the column that carries the string from which the stage will extract the columns.
- **Description**. Allows you to enter an optional description of the input link.

Columns Tab

The entries in the columns grid specify the format of the data read from the input link.



The grid has the standard fields that all column definitions have.

At the least you must define a column to carrying the data string which the stage is splitting. You can also define additional columns if required. Any columns that are defined here and on the Outputs page Columns tab will be passed straight through the stage. So, in the example above, if you defined output columns called othercol1 and othercol2, the data carried by these columns will be passed straight through the stage.

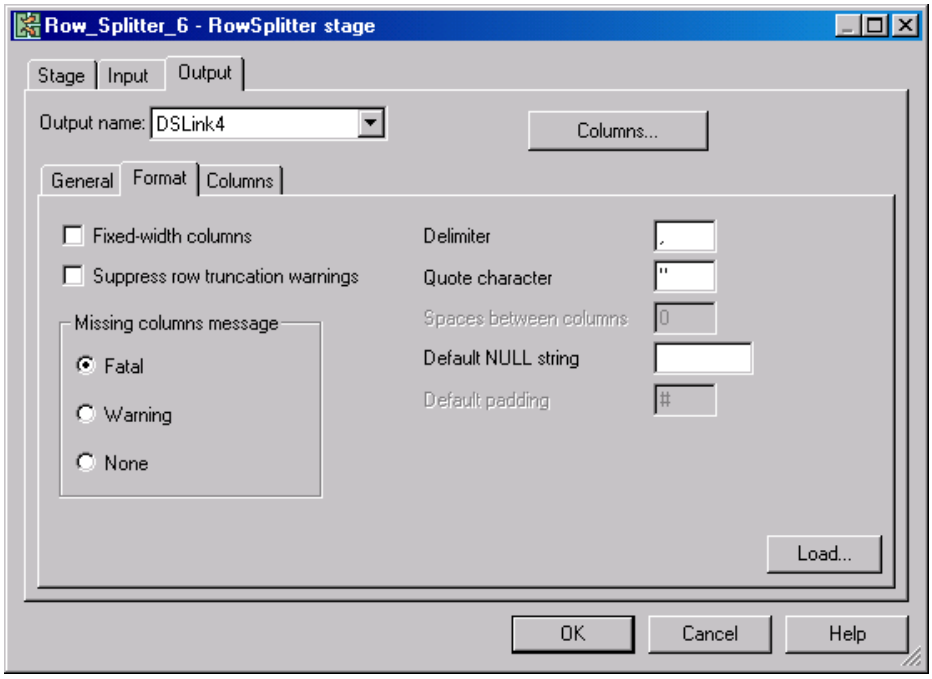
Outputs Page

The **Outputs** page contains various tabs that describe the data being output by Row Splitter.

The **General** tab contains a description field that allows you to enter an optional description of the input link. The **Format** tab and **Columns** tab are described below.

Format Tab

This tab allows you to specify how the input string is formatted, so the stage can split the columns out.



The tab contains the following fields:

- **Fixed-width columns**. Select this check box if the incoming data is in fixed-width format. The width of each field is taken from the SQL display size of the outputs columns (set in the **Display** column in the Columns grid on the **Outputs** page **Columns** tab). This option is cleared by default.
- **Suppress row truncation warnings**. If the input row contains more data fields to be split out into columns than you have defined on the **Columns** tab, you will normally receive warnings about overlong rows when the job is run. If you want to suppress these messages (for example, you may only be interested in the first three columns and happy to ignore the rest), select this checkbox.
- **Missing Columns Message**. If there fewer data fields in the input row than you have defined columns for them to be split into, this option allows you to specify what action to take:
  - Fatal. A fatal error is written to the job log and the job aborts (this is the default.)
  - Warning. A warning message is written to the job log, SQL nulls are writing to the extra columns and the job continues.
  - None. No action is taken. SQL nulls are writing to the extra columns and the job continues.
- **Delimiter**. This option is not available if you have selected **Fixed-width columns**. It specifies the delimiter used to separate the data fields in the input data string. By default this field contains a comma. You can enter a single printable character or a decimal or hexadecimal number to represent the ASCII code for the character you want to use. Valid ASCII codes are in the range 1 to 253. Decimal values 1 through 9 must be preceded with a zero. Hexadecimal values must be prefixed with &h. Enter 000 to suppress the delimiter.

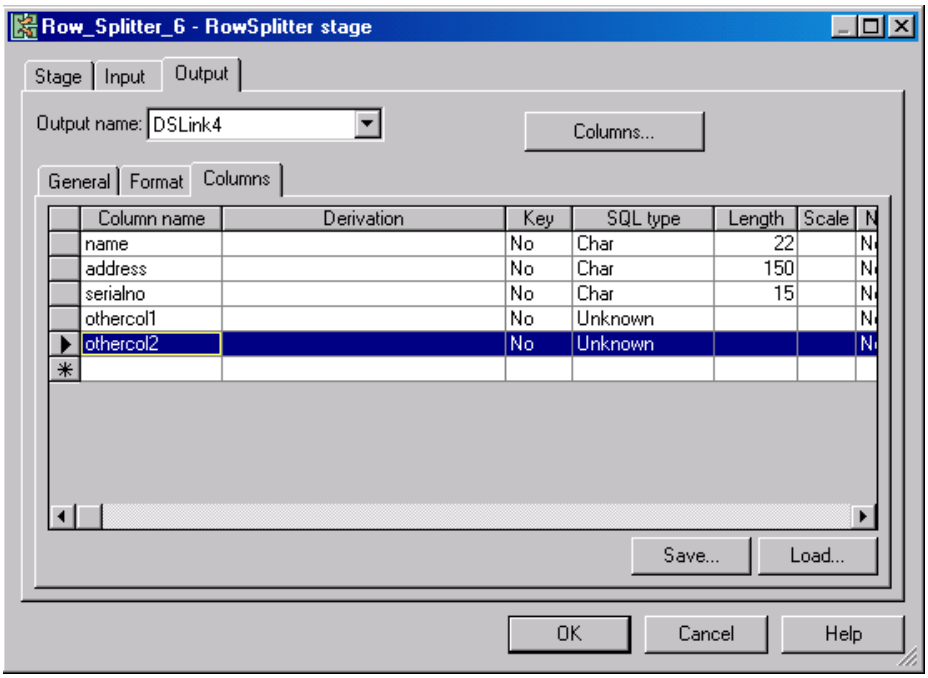
- **Quote Character**. This option is not available if you have selected **Fixed-width columns**. Specifies the character used to enclose strings. By default this field contains a double quotation mark. You can enter a single printable character or a decimal or hexadecimal number to represent the ASCII code for the character you want to use. Valid ASCII codes are in the range 1 to 253. Decimal values 1 through 9 must be preceded with a zero. Hexadecimal values must be prefixed with &h. Enter 000 to suppress the quote character.
- **Spaces between columns**. This option is only available if you have selected **Fixed-width columns**. Contains a number to represent the number of spaces used between columns. By default this is 0.
- **Default NULL string**. Contains characters which, when encountered in an input row, are interpreted as the SQL null value (this can be overridden for individual column definitions in the **Columns** tab).
- **Default padding**. This option is only available if you have selected **Fixed-width columns**. Contains the character used to pad missing columns. This is # by default, but can be set to another character here.

The **Format** tab also has a **Load** button. If you have table definitions that include format information, you can load the format details from these table definitions directly onto the **Format** page:

1. Click **Load**. The **Load Table Definitions** dialog box appears.
2. Browse for the table definition containing the format you want to load.
3. Click **OK**. The format details are loaded.

Columns Tab

The entries in the columns grid specify the format of the data being written to the output link.



The grid has the standard fields that all column definitions have.

The Derivation field is not used.