

Technical Bulletin

Part No. 000-6919

DataStage Command Stage

This technical bulletin describes the Command Stage. This stage executes external programs, routines, and other DataStage jobs within a DataStage job. It also automatically manages the list of output columns.

Copyright © 2003 Ascential Software Corporation
50 Washington Street, Westboro, MA 01581
All rights reserved.

© 2001–2003 Ascential Software Corporation. All rights reserved. Ascential, Ascential Software, DataStage, MetaStage, MetaBroker, and Axielle are trademarks of Ascential Software Corporation or its affiliates and may be registered in the United States or other jurisdictions. Adobe Acrobat is a trademark of Adobe Systems, Inc. Microsoft, Windows, Windows NT, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Ltd. Other marks mentioned are the property of the owners of those marks. This product may contain or utilize third party components subject to the user documentation previously provided by Ascential Software Corporation or contained herein.

Printing History

First Edition (000-6919) for Release 1.0, February 2001
Updated (000-6919) for Release 1.0, August 2002
Updated for Release 1.0, February 2003
Updated for Release 1.0, August 2003

How to Order Technical Documents

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

Documentation Team: Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 1.0 of the DataStage Command Stage Plug-in for DataStage Release 7.0:

- Functionality
- Terminology
- Installation
- Defining the Command Stage
- Defining character set mapping
- Defining input data
- Defining output data
- Using commands
- Stage and link properties

The Command Stage is an active stage that can execute various external commands, including DataStage engine commands, programs, and DataStage jobs from anywhere in the DataStage data flow. You can execute any command, including its arguments, that you can type to the shell of the operating system, such as Windows NT or UNIX. Examples include Perl scripts, DOS batch files, UNIX scripts, and other command-line executable programs that you can call if they are not interactive.

A custom client GUI is available for the Command Stage Plug-in.

You can use the Command Stage anywhere in a job path to invoke an external command. The before- and after-routines that are already available act similarly, except that you can put the Command Stage anywhere in a job stream and call it multiple times in parallel.

If the stage is placed midstream and “Do not forward row data” is cleared, it moves the data to the output link. If the stage is at the end of a path, it executes the command and passes the incoming data through unaltered. The arrival of the row merely causes the command execution.

The Command Stage can have only one input and one output link:

- **Input link.** Specifies a row of actual data or a single row from a previous instance of the Command Stage. You can place the Command Stage midstream or at the end of a job path (with no output link). The stage can have only one input link.
- **Output link.** If you run the Command Stage at the beginning of a job path for an output link, the stage executes the specified command and sends a single row down the output link. Minimally, this row contains the return code from the specified command in the first column. A Transformer stage

can then use DataStage branching operations to process this code. If “Output to link” is selected, the second column holds the output for the command. The stage can have only one output link.

The custom GUI handles the creation of columns on the output link by examining the values of “Output to link,” “Do not forward row data,” and “Do not wait for command.”

For More Information. For information about using DataStage engine, or NLS, see the following table:

If you want information on...	Then see...
Using DataStage BASIC	<i>DataStage BASIC Guide</i>
Using NLS	<i>DataStage NLS Guide</i>
Using MetaStage	<i>MetaStage User's Guide</i>

Functionality

Supported Functionality. The Command Stage has the following functionality:

- More flexibility than using other before- or after-stage routines.
- Visual and textual meta data.
- Graphical invocation of external commands without resorting to job control coding.
- Easier processing of return codes from external commands.
- The stage and its links appear as event meta data within the DataStage Suite Meta Data Management Services.
- Support for NLS (National Language Support). For information, see *DataStage NLS Guide*.
- Support for MetaStage. For information, see *MetaStage User's Guide*.

Unsupported Functionality. The following functionality is not supported:

- Data transformation capabilities on rows flowing through the stage. Use the Transformer and the Aggregator stages to do this.
- Commands requiring user-input or creation of windows. They cause job failures.

- Client access to an RDBMS. If you want to execute an SQL statement, use calls to existing client applications, including DataStage jobs.
- Direct access to DataStage engine commands. You cannot use this stage to return rows that are generated as a result of command execution to the DataStage engine.

Terminology

The following table lists the Command Stage terms used in this document:

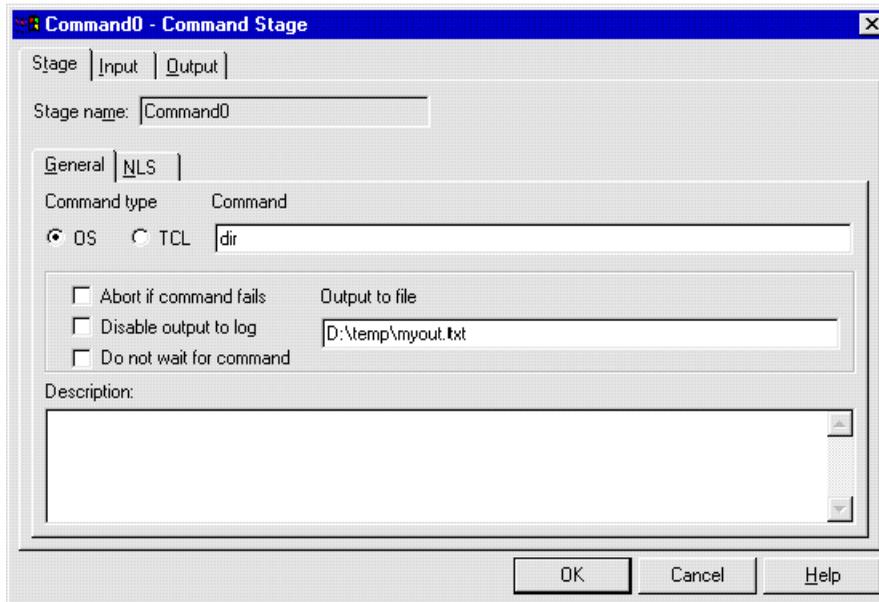
Term	Description
Before and after routines	The external routines that you can define to be called before a job begins and after a job exits. Write these routines in DataStage BASIC. Some stages support before- and after-stage routines. These are called before or after a stage is invoked.
ExecTCL	A built-in DataStage routine that executes DataStage engine commands from a DataStage job.
ExecDOS	A built-in DataStage routine that executes DOS commands from a DataStage job.

Installing the Plug-In

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

Defining the Command Stage

When you use the custom GUI to edit a Command Stage, the **Command Stage** dialog box appears:



This dialog box has the **Stage**, **Input**, and **Output** pages, depending on whether there are inputs to and outputs from the stage:

- **Stage.** This page displays the name of the stage you are editing. The **General** tab defines the command type, the text of the command, the action to take if errors occur, where to write the output, and whether the job waits for the command to complete. You can also describe the purpose of the stage. For details, see “Defining the Command” on page 5.

The **NLS** tab defines a character set map to use with the stage. This tab appears only if you have installed NLS for DataStage. For details, see “Defining Character Set Mapping” on page 6.

- **Input.** This page is displayed only if you have an input link to this stage. It specifies when to execute the command and how to handle the rows from this link.
- **Output.** This page is displayed only if you have an output link to this stage. It specifies how to handle the output from the command.

Defining the Command

The command parameters are set on the **General** tab of the **Stage** page. Specify the appropriate information using the following fields:

- **Command type.** The type of command to be executed. Select one of the following options:
 - **OS.** The stage executes an operating system command.
 - **TCL.** The stage executes a DataStage engine command. You can run DataStage BASIC programs.

For information about using these commands, see “Using Commands” on page 11.

- **Command.** The string to be passed as the command.
- **Abort if command fails.** If selected, the job aborts if errors occur while executing the command.
- **Disable output to log.** If selected, the output from the command is not written to the DataStage log.
- **Do not wait for command.** If selected, the job does not wait for the command to complete before continuing. The job is an independent process and continues to process the data. It executes the command as a thread on Windows NT. The stage waits if the command is still executing after all data is processed.

Selecting this option removes the `COMMAND.RTNCODE` and `COMMAND.OUTPUT` data elements from the output link. Link output is disallowed, but the output and the return code for the command are still written to the DataStage log and the output file. The first column on the output link is not used for the return code.

Additionally, the following options are disabled:

- Abort if command fails (stage)
- Repeat for each row (input)
- Execute command after row (input)
- Do not forward row data (input)
- Output to link (output)
- **Output to file.** Writes output from the command to a file. If you do not specify a pathname, the file is created in the home directory for the project. If you leave the field blank, no output file is created.
- **Description.** Optional. Describe the purpose of the Command Stage.

Defining Character Set Mapping

You can define a character set map for a stage. Do this from the **NLS** tab on the **Stage** page. The **NLS** tab appears only if you have installed NLS.

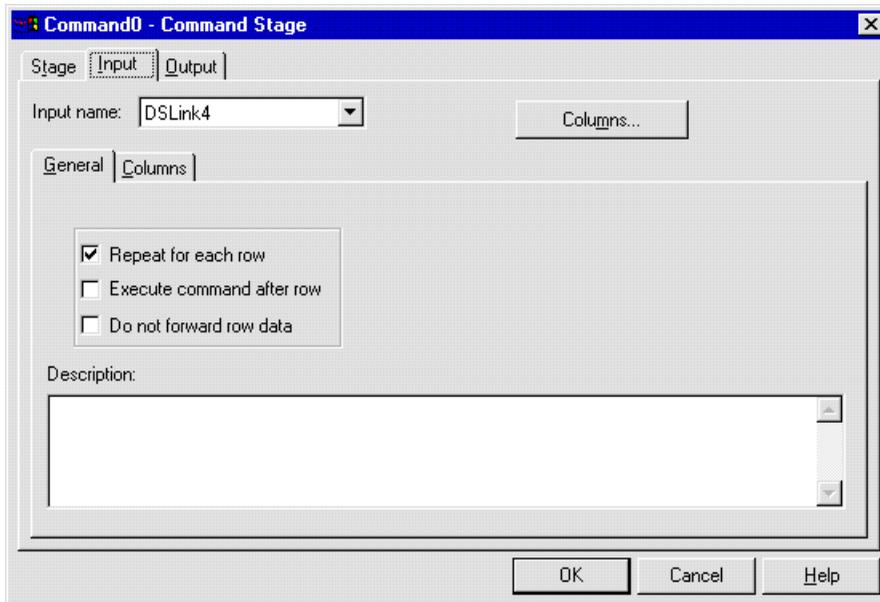
Specify information using the following button and fields:

- **Map name to use with stage.** The default character set map is defined for the project or the job. You can change the map by selecting a map name from the list.
- **Use Job Parameter....** Specifies parameter values for the job. Use the format *#Param#*, where *Param* is the name of the job parameter. The string *#Param#* is replaced by the job parameter when the job is run.
- **Show all maps.** Lists all the maps that are shipped with DataStage.
- **Loaded maps only.** Lists only the maps that are currently loaded.

For more information about NLS or job parameters, see *DataStage Core Developer's Guide* or *DataStage NLS Guide*.

Defining Command Stage Input Data

When a row of actual data or a single row from a previous instance of the Command Stage arrives on an input link of this stage, it executes the specified command. Define the properties of this link and the column definitions of the data on the **Input** page in the **Command Stage** dialog box of the custom GUI.



About the Input Page

The **Input** page has an **Input name** field, the **General** and **Columns** tabs, and the **Columns...** button:

- **Input name.** The name of the input link. Choose the link you want to edit from the **Input name** drop-down list box. This list box displays all the input links to the Command Stage.
- Click the **Columns...** button to display a brief list of the columns designated on the input link. As you enter detailed meta data in the **Columns** tab, you can leave this list displayed.

General Tab

This tab is displayed by default. It contains the following fields:

- **Repeat for each row.** If selected, executes the specified command for each row that arrives on this link. If “Do not wait for command” is selected from the **General** tab of the **Stage** page, this option is disabled to avoid overwhelming the server with processes.
- **Execute command after row.** If selected, executes the specified command after the row is copied and sent to the output link. If there is no output link, this option is disabled. By default, the command is executed asynchronously when the row arrives on the input link.

Selecting this option removes the `COMMAND.RTNCODE` and `COMMAND.OUTPUT` data elements from the output link. Link output is disallowed, but the output and the return code for the command are still written to the DataStage log and the output file. The following options are disabled:

- Do not forward row data (input)
- Output to link (output)
- **Do not forward row data.** If cleared, the stage passes rows through to the same number of columns on the output link, provided it contains both input and output links. You cannot select this option if no output link exists.

If cleared, the column definitions are copied from the input link to the output link. The “Command stage pass thru column” label in the **Description** field identifies each copied column for removal.

- **Description.** Optional. Describe the purpose of the input link.

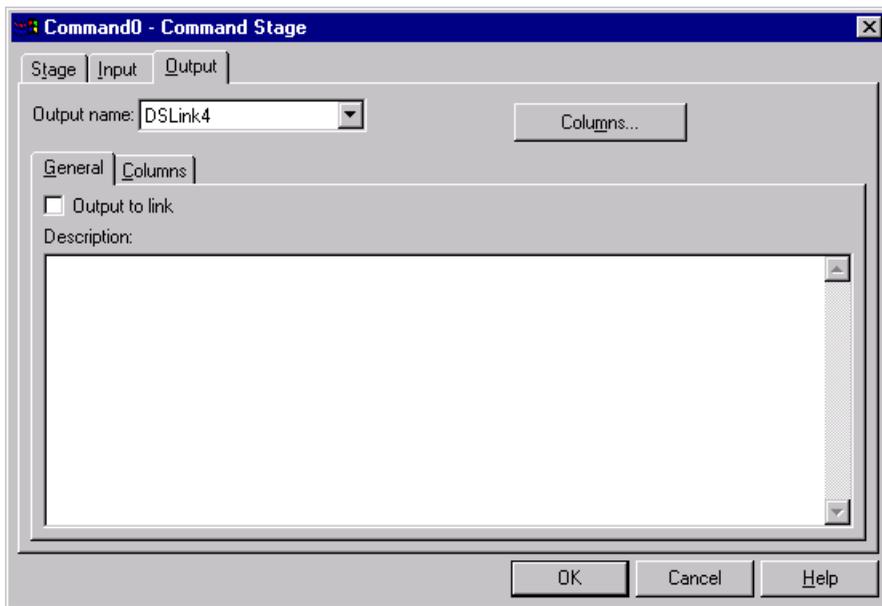
Columns Tab

This tab contains the column definitions for the data written to the data source. The **Columns** tab behaves the same way as the **Columns** tab in the ODBC stage. For a description of how to enter and edit column definitions, see *DataStage Core Developer's Guide*.

Defining Command Stage Output Data

You can write the output of a command as a column on an output link of the Command Stage. The custom GUI automatically manages the output column definitions. The output columns depend more on your choices for field values than on the meta data requirements of their targets. Therefore, you have minimal flexibility in defining Command Stage output columns.

Pass-through columns must have the same data types and sizes as the corresponding input columns. However, you can edit the name, data element, derivation, and description fields for the columns.



About the Output Page

The **Output** page has an **Output name** field, the **General** and **Columns** tabs, and the **Columns...** button.

- **Output name.** The name of the output link. Choose the link you want to edit from the **Output name** drop-down list box. This list box displays all the output links.
- Click the **Columns...** button to display a brief list of the columns designated on this link. As you enter detailed meta data in the **Columns** tab, you can leave this list displayed.

General Tab

This tab is displayed by default. It contains the following fields:

- **Output to link.** If selected, sends the output from the command as the second column on the output link. This `COMMAND.OUTPUT` column holds the output of the command execution.
- **Description.** Optional. Describes the purpose of the output link.

Columns Tab

This tab contains the column definitions for the data being output on the chosen link. The custom GUI automatically manages the output column definitions. For a description of how to enter and edit column definitions, see *DataStage Core Developer's Guide*.

If the “Do not wait for command” option is cleared and the “Output to link” output option is selected, the `COMMAND.RTNCODE` and `COMMAND.OUTPUT` data elements for the column definitions contain the return code and the command output respectively. However, the **Derivation** field is meaningless for this stage.

For details about manually installing these data elements, see “Installing the Plug-In” on page 3.

Using Commands

You can execute any command, including its arguments, that you can type to the shell of the operating system, for example, Perl scripts, DOS batch files, UNIX scripts, and other command-line driven programs that are not interactive or do not request input.

You can set the Command type to TCL on the **General** tab of the **Stage** page to execute DataStage TCL commands and run DataStage BASIC programs.

You can use the *dsjob* command to call other DataStage jobs from the Command Stage. DataStage provides the *dsjob* program to let you run compiled DataStage jobs from a command line instead of from the DataStage Director. *dsjob* has the following simple syntax:

```
dsjob -run [ -mode ] [ -param ] [ -warn ] [ -rows ] [ -wait ]
          [ -stop ] [ -jobstatus ] [ -userstatus ] project job
```

For full syntax information, see *DataStage Server Job Developer's Guide*.

Note: If you select **Omit** in the **Attach to Project** dialog box as you start DataStage, you must use the `-user` and `-password` options when you use the *dsjob* command.

You can write the output of a command to any of the following:

- The DataStage log. This is the default.
- Output links. If output links exist, you can write the output as a column on the link, in addition to the return code for the command. The return code is automatically sent as the first column on the output link.
- A file.

Note: Since the stage sends the return code for the command as the first column of an output link, you must manually add the mandatory columns to the column definitions for the output link if you use the standard grid editor. The custom GUI handles this automatically.

