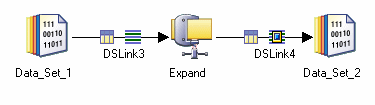
The Expand stage is a processing stage that converts a previously compressed data set back into a sequence of records from a stream of raw binary data.

The Expand stage is a processing stage. It can have a single input link and a single output link.

The Expand stage uses the UNIX uncompress or GZIP utility to expand a data set. It converts a previously compressed data set back into a sequence of records from a stream of raw binary data. The complement to the Expand stage is the Compress stage which is described in [Compress stage](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Compress_Stage.html?view=kc).



The stage editor has three pages:

* [**Stage Page**](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Stage_Page_expand_stage.html?view=kc). This is always present and is used to specify general information about the stage.
* [**Input Page**](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Input_Page_expand_stage.html?view=kc). This is where you specify details about the data set being expanded.
* [**Output Page**](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Output_Page_expand_stage.html?view=kc). This is where you specify details about the expanded data being output from the stage.

This section specifies the minimum steps to take to get an Expand stage functioning.

**About this task**

InfoSphere® DataStage® has many defaults which means that it can be very easy to include Expand stages in a job. InfoSphere DataStage provides a versatile user interface, and there are many shortcuts to achieving a particular end, this section describes the basic method, you will learn where the shortcuts are when you get familiar with the product.

To use an Expand stage:

* In the Stage page [**Properties Tab**](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/r_deeref_Properties_Tab_expand_stage.html?view=kc) choose the uncompress command to use. This is *uncompress* by default but you can also choose *gzip*.
* Ensure column meta data is defined for both the input and output link.

The General tab allows you to specify an optional description of the stage. The Properties tab lets you specify what the stage does. The Advanced tab allows you to specify how the stage executes.

Use the Properties tab to specify how the Expand stage operates.

The **Properties** tab allows you to specify properties which determine what the stage actually does. The stage only has a single property which determines whether the stage uses uncompress or GZIP.

| **Category/Property** | **Values** | **Default** | **Mandatory?** | **Repeats?** | **Dependent of** |
| --- | --- | --- | --- | --- | --- |
| Options/[Command](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/r_deeref_Options_Category_expand_stage.html?view=kc) | uncompress/gzip | uncompress | Y | N | N/A |
| *Table 1. Properties* | | | | | |

Use the Options category to specify how the Expand stage operates.

## Command

Specifies whether the stage will use uncompress (the default) or GZIP.

This tab allows you to specify options.

This tab allows you to specify the following:

* **Execution Mode**. The stage can execute in parallel mode or sequential mode. In parallel mode the input data is processed by the available nodes as specified in the Configuration file, and by any node constraints specified on the Advanced tab. In Sequential mode the entire data set is processed by the conductor node.
* **Combinability mode**. This is Auto by default, which allows InfoSphere® DataStage® to combine the operators that underlie parallel stages so that they run in the same process if it is sensible for this type of stage.
* **Preserve partitioning**. This is **Propagate** by default. The stage has a mandatory partitioning method of **Same**, this overrides the preserve partitioning flag and so the partitioning of the incoming data is always preserved.
* **Node pool and resource constraints**. Select this option to constrain parallel execution to the node pool or pools or resource pool or pools specified in the grid. The grid allows you to make choices from drop down lists populated from the Configuration file.
* **Node map constraint**. Select this option to constrain parallel execution to the nodes in a defined node map. You can define a node map by typing node numbers into the text box or by clicking the browse button to open the Available Nodes dialog box and selecting nodes from there. You are effectively defining a new node pool for this stage (in addition to any node pools defined in the Configuration file).

**Note**In the **Node map constraint** text box, you can enter jobs parameters as well as numbers. You can enter a single parameter, for example #testnode#, or you can enter a comma separated lists of parameters, for example #testnode#, #testnode2#. The browse button next to the text box will display a list of the node names from the last configuration file that was referenced by the job, but the browse button will not display the node names that were specified by the job parameters.

The Input page allows you to specify details about the data set being expanded. There is only one input link.

The Input page allows you to specify details about the data set being expanded. There is only one input link.

The General tab allows you to specify an optional description of the link. The Partitioning tab allows you to specify how incoming data on the source data set link is partitioned. The Columns tab specifies the column definitions of incoming data. The **Advanced** tab allows you to change the default buffering settings for the input link.

Details about Expand stage partitioning are given in the following section. See ["Stage Editors,"](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Stage_Editors.html?view=kc) for a general description of the other tabs.

The Partitioning tab allows you to specify details about how the incoming data is partitioned or collected before the expansion is performed.

By default the stage uses the **Same** partitioning method and this cannot be altered. This preserves the partitioning already in place.

If the Expand stage is set to execute in sequential mode, but the preceding stage is executing in parallel, then you can set a collection method from the **Collector type** drop-down list. This will override the default auto collection method.

The following Collection methods are available:

* **(Auto)**. This is the default collection method for the Expand stage. Normally, when you are using Auto mode, InfoSphere® DataStage® will eagerly read any row from any input partition as it becomes available.
* **Ordered**. Reads all records from the first partition, then all records from the second partition, and so on.
* **Round Robin**. Reads a record from the first input partition, then from the second partition, and so on. After reaching the last partition, the operator starts over.
* **Sort Merge**. Reads records in an order based on one or more columns of the record. This requires you to select a collecting key column from the **Available** list.

The Partitioning tab normally also allows you to specify that data arriving on the input link should be sorted before the expansion is performed. This facility is not available on the expand stage.

The Output page allows you to specify details about data output from the Expand stage

The Expand stage only has one output link.

The General tab allows you to specify an optional description of the output link. The Columns tab specifies the column definitions of the data. The Advanced tab allows you to change the default buffering settings for the output link.

See ["Stage Editors,"](https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.parjob.dev.doc/topics/c_deeref_Stage_Editors.html?view=kc) for a general description of the tabs.