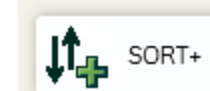


Sort Plus

Overview



The **Sort Plus Component** is an SSIS Data Flow asynchronous transformation component, which is capable of sorting very large amounts of data. It uses a sophisticated proprietary algorithm that employs smart chunking and caching of data on disk. It enables an optimal utilization of your server's resources, by doing all processing in parallel, using a configurable number of threads.

With smaller amounts of data (e.g. well below 1GB), its speed is similar to the speed of the built-in Sort component. However, unlike the built-in **Sort** component which usually hangs when trying to process big volumes of data (e.g. more than 2 GB), our Sort Plus component completes its job and does so in a reasonable time.

Unlike the built-in **Sort** transformation, **Sort Plus** supports multiple inputs. The first input is used as a basis for selecting output columns and sorting keys. The others must have the same structure (i.e. columns with the same names and data types). For optimal throughput, the data from all inputs is processed in parallel.

Additionally, unlike the built-in component, large text columns (i.e. *DT_TEXT* and *DT_NTEXT*) can be part of the sorting key. Please note, that in such scenarios, you should expect decreased performance compared to using a fixed size string column (i.e. *DT_STR* or *DT_WSTR*).

Aside, from being able to process much larger amounts of data, the functionality of COZYROC's Sort Plus component is very similar to the built-in one.

Demonstration



Now Playing

Basics

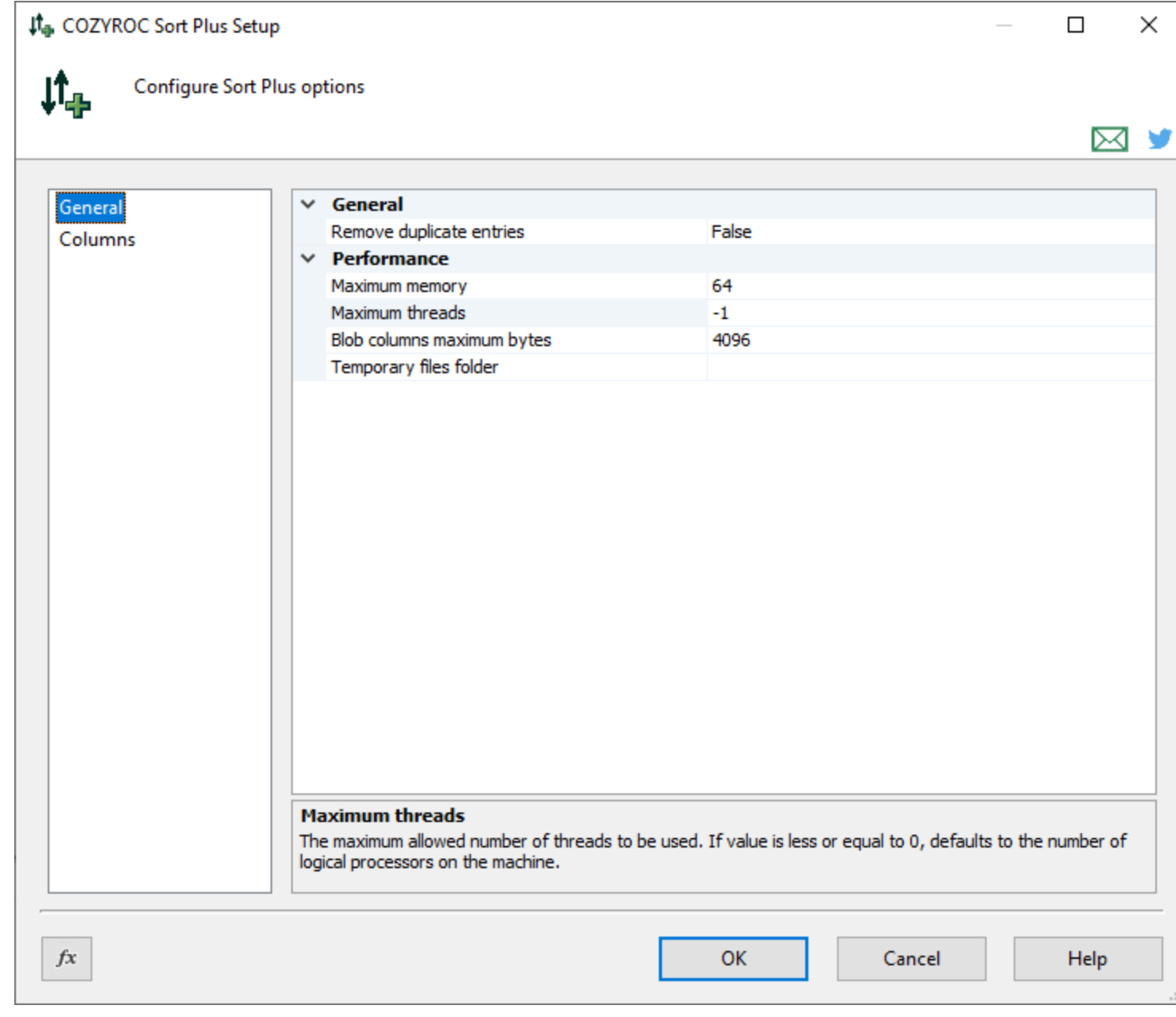
6:40

UP NEXT

Handling multiple inputs

Parameters

General



- Remove duplicate entries**

Setting this option to *true* would output only one row for all rows that have equal sorting keys. As the processing is parallelized, this row wouldn't necessarily be the first encountered one. The default is *false*.
- Maximum memory**

Use this parameter to specify the maximum amount of memory (in Megabytes) that the component is allowed to use. The default is *64MB*. Please note, that as the size of the input rows can vary depending on the data, there might be some small fluctuations in the actual memory usage, so this parameter will be used as a guidance and not a strict limit.
- Maximum threads**

Use this parameter to specify the maximum number of concurrently running sorting sub-tasks. Each chunk of data (i.e. pipeline buffer) gets processed in separate thread that performs a calculation-heavy sub-task. Leaving the default value *-1* would instruct Sort Plus to use the number of *logical threads* (i.e. the number of cores multiplied by the number of hyperthreads per core available). Setting the parameter value to a number higher than the number of logical threads is not recommended.
- Blob columns maximum bytes**

Use this parameter to specify the maximum bytes expected from an input BLOB column to hold. Such columns can hold large amounts of data in theory, but the limit is rarely reached in practice. In order to manage memory allocation in optimized way, try to set an accurate value for this parameter. The default value is *4096* bytes.
- Temporary files folder**

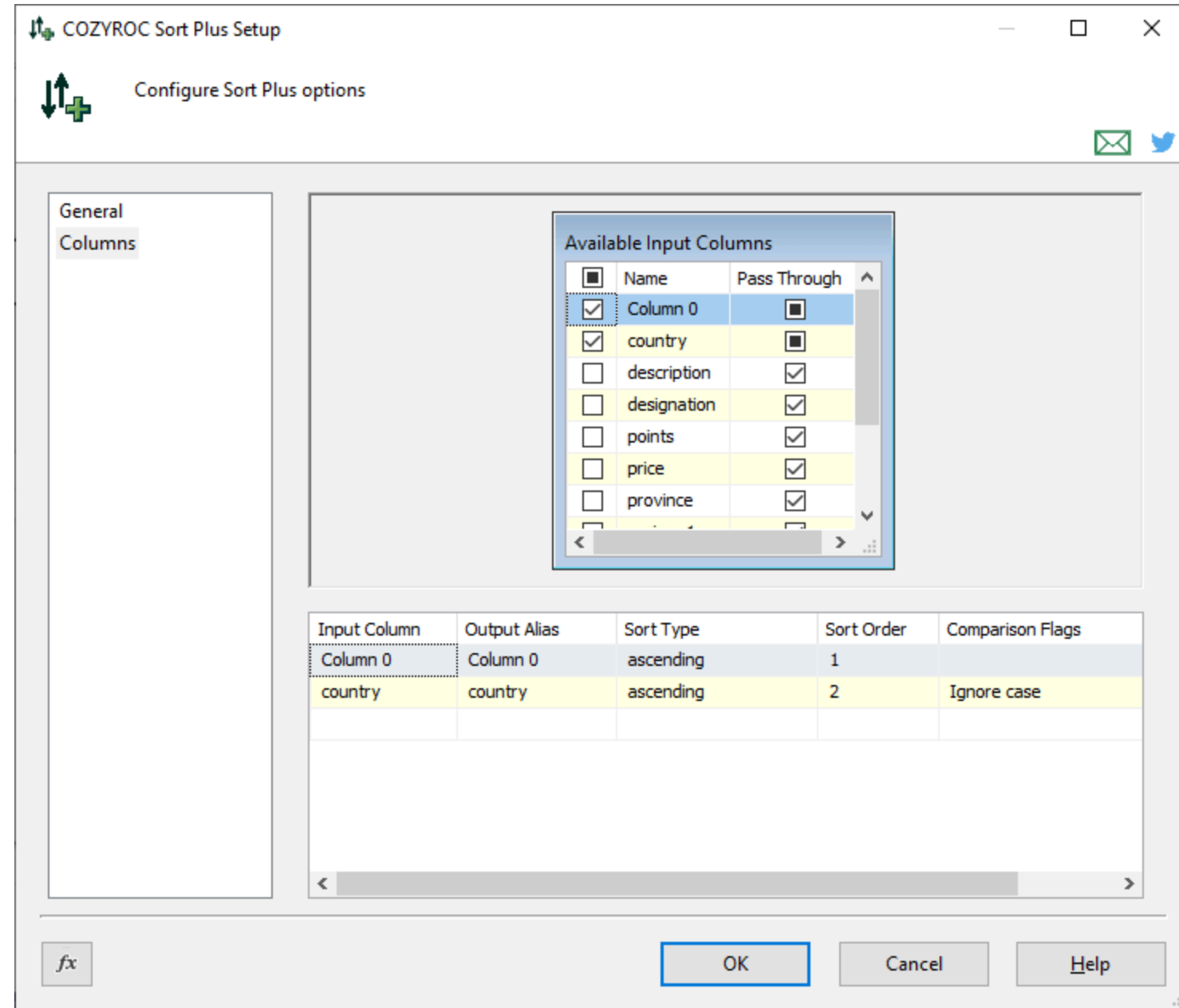
Dealing with the data in chunks and offloading already sorted ones to disk is a key functionality of *Sort Plus*. Use this parameter to specify a folder where you want temporary (cache) files to be stored. If left empty, the system's default temporary directory will be used.
- LoggingLevel**

Use this parameter to specify logging level. When enabled the component will write information messages on key sorting events. Along with each event are logged several parameters of the used memory. Available levels - *Off, Basic, Verbose*. *Basic* level logs only the basic events, *Verbose* level logs all events.
- MaxParallelOffloads**

Use this parameter to specify maximum allowed number of threads to be used for offload of sorted rows to a temporary file. The value is taken into account when calculating the maximum number of rows in memory. Increasing the value will decrease the rows in memory. Usually *2* parallel offloads is a good setting. Leaving the default value *-1* would instruct Sort Plus to use *2* threads. If you specify too big of the value, the property value at runtime will be up to half of the threads, specified in *MaxThread*.

Columns

The columns / sorting key setup page is quite similar to the one used by the built-in *Sort* component.



Each column can be selected as part of the sorting key and/or (deselected) to be included in the output. The latter is controlled with the *Pass Through* checkbox. Of course, all columns that are part of the sorting key **must** be sent to the output. The default is to pass through all of the columns.

The sorting key configuration table allows you to: change the sorting key position (**Sort Order**) for each column (**Input Column**), set a different name for it (**Output Alias**) and specify the direction of sorting (**Sort Type**). Additionally, if any of the selected sort columns are of string/text type, the **Comparison Flags** may be useful.

Outputs

The component has a single output which, by default, would contain all the columns. Any columns deselected in the *Pass Through* configuration, will be excluded in the output. The names of the output columns can be modified using *Output Alias*.

Knowledge Base

- Where can I find the documentation for the Sort Plus? [See more...](#)

What's New

- 2.0 SR-2**
 - New: A new parameter **Blob columns maximum bytes** to configure maximum expected BLOB column size.
 - New: A new parameter **LoggingLevel** to provide more detailed tracing information about the sort process at runtime.
 - New: A new parameter **MaxParallelOffloads** to configure number of threads used for input data off-loading.
 - Fixed: Component consumed more memory than specified in **Maximum memory** parameter.
- 1.9 SR-2**
 - Fixed: Crash with error "System.ArgumentException at System.Collections.Generic.IntrospectiveSortUtilities.ThrowOrIgnoreBadComparer..." (Thank you, John).
- 1.9**
 - New: Introduced component.

Related documentation

- Database Destination
- Database Partitions Task
- Database Source
- DB2 Destination
- Informix Destination
- ODBC Destination
- Oracle Destination
- Script Oracle Destination
- Snowflake Connection
- Sybase Destination
- Table Difference
- Teradata
- Teradata Destination
- Teradata Source

COZYROC SSIS+ Components Suite is free for testing in your development environment. [Download Now!](#)

A licensed version can be deployed on-premises, on Azure-SSIS IR and on COZYROC Cloud.

Newsletter

Join VIP subscribers getting **Exclusive** COZYROC SSIS+ Library updates, private Q&A and more.

No spam, ever. Just great stuff!

[Join](#)

Contact Us

Address: 4350 Lassiter at North Hills Ave, Raleigh, NC 27609, United States
Phone: +1 (919) 249-7421
Pre-Sales: Schedule a call
Support: Help Center

Support

- Help Center
- Getting Started
- Licensing
- Tutorials
- Forum

Follow Us

