# KNIME Cluster Execution

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNIME Cluster Execution</td>
<td>1</td>
</tr>
<tr>
<td>INSTALLATION</td>
<td>2</td>
</tr>
<tr>
<td>Configure the SGE Cluster Execution plug-in via its preference page</td>
<td>3</td>
</tr>
<tr>
<td>Quickstart Guide to KNIME Cluster Execution</td>
<td>5</td>
</tr>
<tr>
<td>Cluster execution overview</td>
<td>5</td>
</tr>
<tr>
<td>Split jobs</td>
<td>5</td>
</tr>
<tr>
<td>Disconnecting from an executing node</td>
<td>10</td>
</tr>
<tr>
<td>Cluster job log file views</td>
<td>10</td>
</tr>
<tr>
<td>Temporary Files</td>
<td>11</td>
</tr>
<tr>
<td>Pre- and Post-scripts</td>
<td>12</td>
</tr>
</tbody>
</table>
**INSTALLATION**

SGE Cluster Execution is available for **Linux only** and for KNIME versions 2.1 upwards. Use the Update Wizard to install the plug-in.

KNIME should be installed in a location that is accessible by all SGE cluster machines and by the user's Linux machine (although it is possible to have separate installations for the client and the cluster).

The user's KNIME installation must contain the SGE plug-in. If the cluster job should re-submit jobs to the cluster, this installation requires the plug-in as well.

All machines of users that should be allowed to submit KNIME jobs into the cluster must be registered as submit hosts with the SGE master.

Users must be able to submit jobs directly (i.e. must be able to submit from the command line with the native `qsub` command). Custom submit scripts mandatory for job submissions are not supported.

The `$SGE_ROOT/lib/<arch>` is added to the user's `LD_LIBRARY_PATH` shell variable. Alternatively add the following line to the “knime.ini” file in all installations:

```
-Djava.library.path=<SGE_ROOT>/lib/<arch>
```

where `<arch>` is usually either “lx24-amd64” in a 64-bit installation or “lx24-x86” for 32-bit installations.
Configure the SGE Cluster Execution plug-in
(If the preference page is not available either the $SGE_ROOT variable is not set properly or the SGE libraries are not in the LD_LIB path (and not specified in the ini-file.)

These are the default settings for all nodes executed on the cluster. They can be overwritten in each node's dialog.

- **KNIME executable**: Specify the full path to the KNIME executable as seen from the cluster computer node (the execution host). You can include commands or variables to select the right executable, depending on the architecture of the cluster node. This line appears in the shell script that is executed on the cluster node. The architecture selection must work in the selected shell (see below).

- **KNIME arguments**: You can specify command line arguments that are passed to the KNIME instance launched on the cluster node. Entering arguments here causes the “knime.ini” file (with all the arguments defined in it) to be ignored.

- **Native SGE arguments**: Specify any “qsub” arguments here. E.g. to send jobs to a specific queue, ...

- **Location of the shared dir**: This is where the client (submitter) stores the
temporary workflows. This directory must be mounted on the client machine as well as on all cluster node computers. Specify here the path as seen from the client (local) computer.

- **Shared dir (cluster node path):** If the mount point of the shared directory is different on the cluster nodes, specify this here – otherwise leave it empty.

- **Export preferences:** If the nodes submitted to the cluster need special preferences that are set in the submit client, set a check mark here. Otherwise the default preferences in the cluster installation are used.

- **bash/tcsh:** Select the type of shell that is your preferred and default login shell. The script generated for each cluster job is created for this kind of shell. All variables and/or commands used in the settings above must work in the selected shell.

- **Delete temp files policy:** Select whether the temporary files in the shared directory should either always be deleted, or only if the cluster job succeeds, or never.
Quickstart Guide

**Prerequisites:**

- **Linux only**: Node submission works from Linux computers only.
- **Shared Directory**: A shared directory must be accessible from the submit host (your Linux machine) and all cluster node machines. Temporary workflows are stored there for cluster submission.
- **SGE installation**: SGE must be installed on your Linux machine. Your computer must be registered as submit host with the SGE master. The “$SGE_ROOT” variable must be set properly.

**Cluster execution overview**

If the SGE Cluster Execution plug-in is installed, nodes, meta-nodes, and workflows can be executed on a cluster node. (In the following we always refer to nodes, but this term also applies to meta-nodes and workflows respectively, unless noted otherwise.)

The cluster support also allows the KNIME user to disconnect from running nodes. The local KNIME workflow can be closed while the cluster job continues to run. When the workflow is opened again, KNIME reconnects to the cluster.

Cluster execution is provided for all KNIME nodes. Nodes do not have to be designed in a special way. All nodes from all vendors should be ready for grid execution.

There is a certain overhead related to cluster submission. Only time-consuming jobs should be sent to the cluster, to ensure that the overhead of creating the job, submitting it and copying the results is only a small portion of the job’s overall computation time.

**Split jobs**

There are two ways of executing a node on a cluster: either “normally” as one job, or as multiple jobs on chunked input data.
Normal execution behaves as if the node had been executed locally - and is always the safe option.

Chunked execution splits the input table into partitions and creates a cluster job for each partition. Each job contains a copy of the node and receives one partition as input. All copies are executed and the results of all “split-jobs” are joined back together to one output table – which is subsequently implanted in the original workflow.

**Settings for Cluster Execution:**
If the SGE cluster plug-in is installed, the KNIME preferences include an additional page and each node dialog has an additional tab containing the controls to enable cluster execution for the respective node and to adjust its settings.

**SGE Job Manager Preference page:**
The default settings can be adjusted in the SGE preference page; to do so, click the menu “File” → “Preferences...”. Open the “KNIME” section and select the SGE Job Manager Settings page. The settings are explained above. These default settings can be overwritten for each node's dialog.

**Node specific SGE Job Manager settings:**
To select cluster submission for a node, click the “Job Manager Selection” tab and select the “SGE Shared Dir” job manager. (The “<<default>>” job manager executes the node locally.)

The “Split Execution” tab, determines the way the node is executed on the cluster (normal or split):
Split input table and execute node on chunks

If you do not select the setting “Split the input table”, normal node operation is performed on the cluster.

If “Split input table” is checked, the input table of the node is split into partitions and multiple copies of the node are submitted – each operating on the partial input data only. After the nodes finish on the cluster their result tables are joined to form one output table for the local node. (Please read the notes and warnings below before enabling this option.) This option is disabled for meta-nodes and workflows.

Split port selection:

Only one input table is split, all other tables are presented to the cluster jobs in full length. In the dialog you can specify the port index whose table should be split.

Number of split jobs:

The number of partitions can be adjusted in this section. It also determines the number of jobs submitted for this node. If you enter a fixed number of jobs, the size of the input table to each split job (and its time to finish) depends on the size
of the node's input table. Alternatively you can specify the number of rows in the
input table of each cluster job. Here, the number of jobs created depends on the
size of the node's input table (the number of actually used rows could differ if the
number of input rows cannot be divided by the specified number).

**Append index to result row IDs:**
Check this, if the node creates new row IDs in its output table. When the result
tables of the split jobs are joined (appended), the row IDs of all rows must be
unique within the result table. If the node generates new row IDs (i.e. independent
and unrelated to the input table row IDs) the IDs of all result tables will be the
same (because each node copy starts at row zero). When this option is checked, the
chunk index is appended to the row IDs created by each chunk, making the ID
unique.

**Note and warning about split execution**
All nodes (except meta-nodes and workflows) can be set for split execution so that the
cluster job performs exclusively on a part of the input table.

**This only leads to the expected and correct results, if the node performs a
“row-wise” computation.** This means that the result of the node must be the
same if run on the entire table or only on part of it.

This can only be the case, when the node adds one (or more) column(s) to each row in
which the computed result is stored. Or if it creates new rows where each new row is
clearly related to an input row and contains the result related to that one specific
input row.

It is still possible that the result is not correct (even though the nodes perform row-
based computations) and it might still be difficult to determine if the results are
correct. If the node, for example, computes the average of some column before it
starts computing its results, this average will be different if it only operates on a
partial table. It is therefore important to clearly understand what a node does, before
executing it in chunks.
Overwriting default SGE Cluster Execution settings:
In the “KNIME Executable” tab, you can overwrite the settings from the SGE preference page (for example, if the node needs to be queued into a specific SGE queue, because it requires a special license):

To remove the default values, deselect the checkbox “Use default values”. If, however, you want to use some of the default values, click the “Set to defaults” button and the values are re-entered into the fields. You can now modify the values you want to change. See the description above for information on the values to enter in each field.

Additional column “SplitJobResult”:
The output table of a node that was executed as a split job on the cluster contains an additional column “SplitJobResult” created by the SGE Job Manager. There are two possible values in this column: “success” and “failed”.
If some of the split-jobs of a chunked node fail, KNIME still collects the results from the succeeding jobs, creates the joined output table and marks the node as executed.
On output port #0 (only!) introduces “fake” rows for each failed chunk. These “fake” rows have the same row IDs as the corresponding rows of the input partition to that failed job. Also, the size of the inserted fake part is the same as the input partition.
The content of the inserted fake rows is taken over from the input partition, but only if the name and the type of the columns match. In new columns a missing value is inserted.

In order to distinguish the inserted “fake” rows from the rows of a succeeding job, the additional column is added to the output table of each split job. In each row of a succeeding job it contains “success”, and in “fake” rows of a failing job it contains “failed”.

**Disconnecting from an executing node**
If a node is submitted for cluster execution you can disconnect the client, re-connect to the job later and collect the results. After the node's job (or all of its split jobs) has been inserted into an SGE queue, the workflow can be closed (after saving it). KNIME could be terminated. If the workflow with the executing node is opened again, it tries to reconnect to the SGE cluster and retrieves the status of the node's job(s). If the jobs are finished, it immediately starts reading the results into the client. Partially-finished jobs can no longer be disconnected. After the result has been fully applied to the node, you must save the workflow to preserve the result.

**Cluster job log file views**
Nodes set for cluster execution have an additional node view that shows the log files from the last cluster execution run. To open this view, select the “Job Manager View” from the node's context menu.

The view contains three log files (if they exist):

The “KNIME log file” tab shows the KNIME messages from the execution of the node.
In the “Standard Output” tab the lines are displayed that were sent to standard output by the cluster job.

The output of the cluster job to standard error is shown in “Error output”.

If the node was split into multiple jobs, you can select from which split job you want to see the output.

These log files are preserved when the workflow is saved. Other log files (from external tools for example) are not available and not saved.

**Temporary Files**

For each node that is executed on the SGE cluster a temporary mini-workflow is created (containing a copy of the node to execute), which is subsequently executed by a new KNIME instance (the “batch executor”) on the cluster machine.

This workflow and its accompanying files are stored under the shared directory specified in SGE preferences. The SGE job manager creates a directory called “knimeJob_<date>_<id>”. With <id> being any number making the name unique.

This job directory contains the following content:

- **InvokeKNIME.sh** script, which is the command submitted to the SGE cluster.
- This script starts KNIME and executes the mini-workflow.
**grid_job_workflow** contains the mini-workflow that is run in the cluster KNIME instance.

**runtimeWS** is a directory used as workspace by the cluster KNIME instance. It is basically empty – and only contains the KNIME log file (as well as some more configuration-specific files).

**preferences.epf** only exists if the preferences are exported from the submit client to the cluster job – and subsequently contains these preferences.

**stdout.txt** and **stderr.txt** contain the output of the cluster job to standard output and error output respectively.

This job directory plus content is deleted after the cluster job finishes, according to the deletion policy set in the SGE preference page.

**Pre- and Post-scripts**

The SGE Job Manager creates a script (InvokeKNIME.sh) that is submitted as a command to the SGE cluster. This script starts KNIME and executes the node. (See section on “Temporary Files” above.) This script sources other scripts at specific locations (if they exist), allowing for certain commands to be executed before and after KNIME is launched. These other scripts must be stored in a directory called “knimeControl” in the shared directory. The script that is called before KNIME is launched has to be named “preKNIME”; the script that is called after KNIME has finished is named “postKNIME”. By default these files do not exist.

**Unable to load the required library 'drmaalib':**

If you see this message at the beginning of a session, it means that KNIME was not able to find the SGE installation and the required SGE libraries.

Please make sure your SGE_ROOT variable is set in the shell. (“echo $SGE_ROOT”). Make sure KNIME finds the lib directory in the SGE installation: Either specify the LD_LIBRARY_PATH variable or add a line to each “knime.ini” file (located in the KNIME installation directories). The line added would look like this “-Djava.library.path=/<sge_root>/lib/lx24-<arch>” depending on your SGE installation directory and the KNIME binary (64-bit or 32-bit).