

## Running KNIME Workflows in a Compute Cluster Environment

Compute clusters often run idle because of a lack of applications that can be run in a cluster environment and due to the enormous effort required to operate, maintain, and support applications on the grid. KNIME Cluster Execution tackles this problem by providing a thin connection layer between KNIME and the cluster. This allows every node running in KNIME and every application integrated in KNIME to be executed on the cluster. Submission of data to the cluster and collection of the results is made very simple. Long-running analysis workflows can be executed on the compute cluster, freeing up local resources for other productive work. Expensive licenses can be shared by running the referring nodes on the compute cluster, reducing costs.

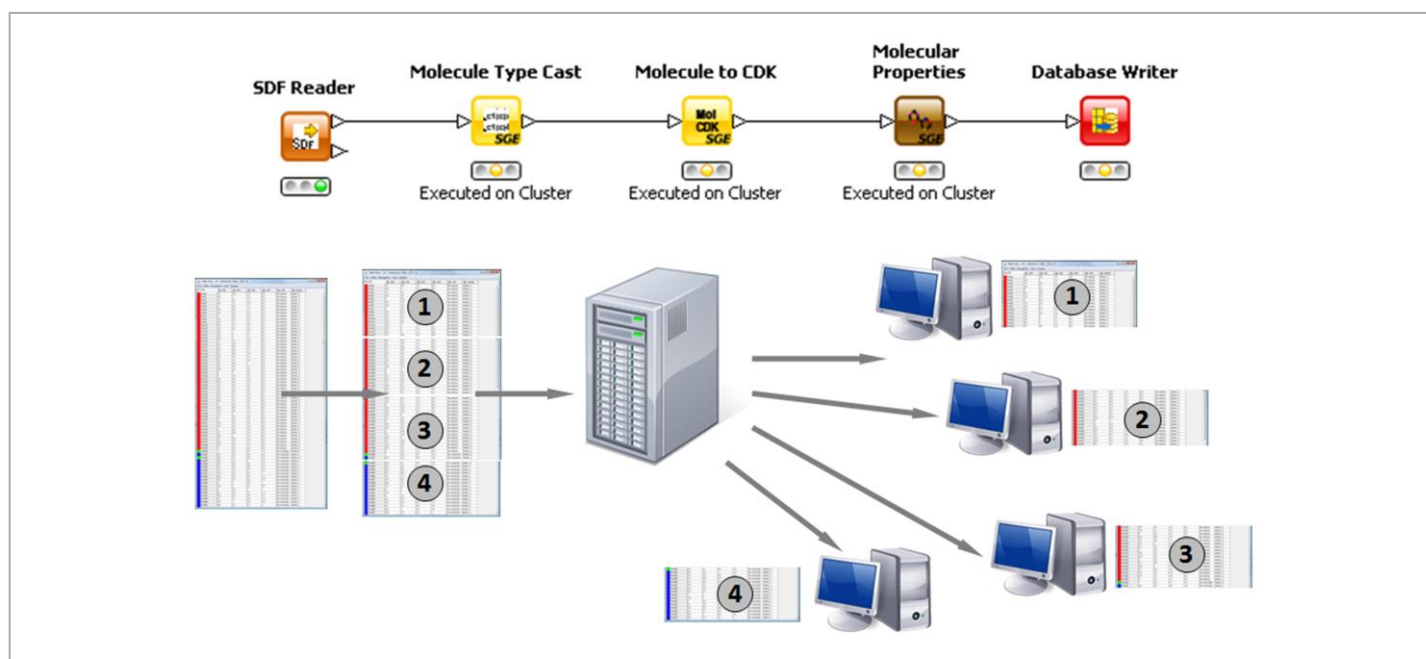


Figure 1: Compute-intensive tasks are split into subsets and executed on different resources in the cluster

### Cluster Execution Concept

Workflows are created in the familiar manner using the KNIME Desktop. Now the user can make selections for each individual node to specify whether it should be executed locally on the PC or submitted to the cluster. A preference setting in KNIME specifies which cluster engine is used. Node-specific settings can also be entered to define how the cluster is used.

The data that will be processed can be split into multiple subsets (Figure 1), which are subsequently sent to the master server in the cluster. The master coordinates execution of the task. In split workflows parallel nodes can be executed on separate resources in the cluster (Figure 2). Once the remote execution of all tasks is finished KNIME collects the results from the remotely-executed nodes and completes the workflow execution locally.

### Applications

Workflows that process a large number of data sets or that require CPU-intensive calculations are ideal candidates for using KNIME Cluster Execution.

The solution is already successfully used for 3-D modeling calculations and virtual docking experiments in pharmaceutical research.

Another area of use is the analysis and mining of huge data volumes found in CRM and telecom databases, or financial systems.

KNIME Cluster Execution is used in these industries by KNIME power users to provide calculation results in a reasonable time frame, which would be too time consuming with just regular desktop execution.

This enables the KNIME user to provide mission critical results to business processes and management.

### KNIME Integration

KNIME Cluster Execution is an extension of KNIME Desktop. But it can be easily integrated with KNIME Team Space for advanced use of shared metanodes that can be submitted to the compute cluster as well.

The KNIME Server can also be combined with KNIME Cluster Execution to start calculation intensive tasks on the server where the workflow containing cluster enabled nodes can call out to the compute cluster. On workflow completion the results can be downloaded from the server or full reports can be saved in the desired report document format like PDF, Word, Excel or PowerPoint.

In combination with KNIME Web Portal, KNIME Server the submission of parameters and data files for the workflow through a simple web browser interface.

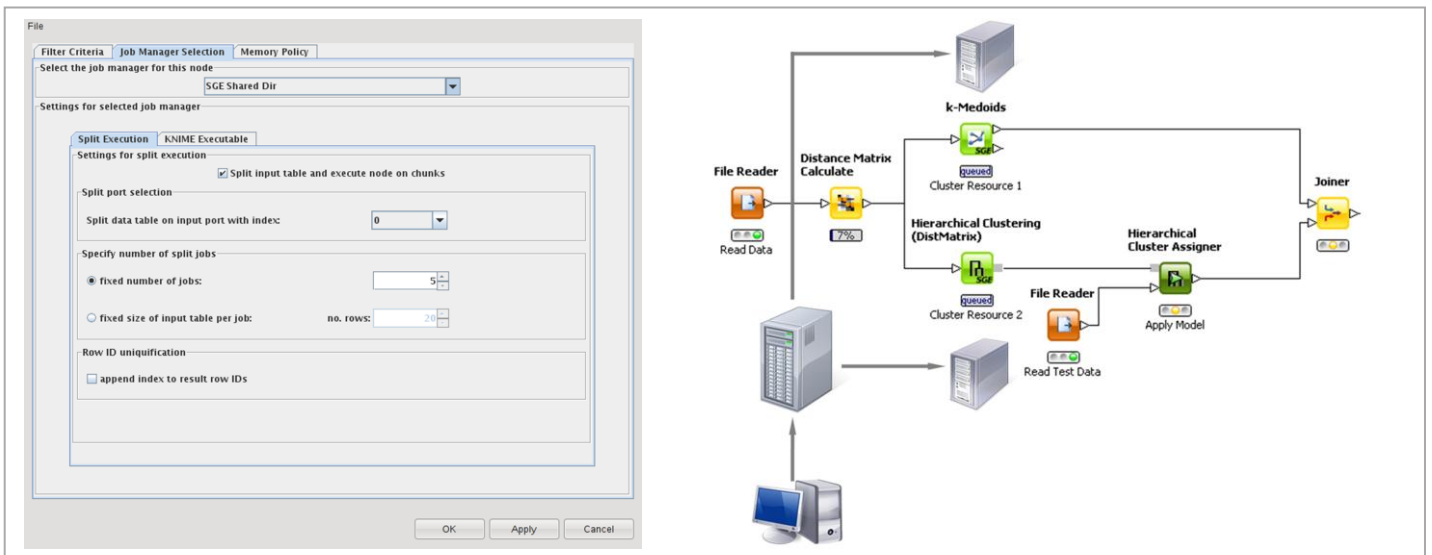


Figure 2: UI for KNIME Cluster Execution and executing individual nodes on different resources in the cluster environment

### Advantages

An important benefit is the gain in performance of calculation intensive workflows. KNIME Cluster Execution enables you to disconnect from running jobs to continue work on other urgent tasks and later reconnect to those jobs to check for status changes and retrieve the results. Third party nodes can also be routed to dedicated servers, making it possible to distribute software which usually does not offer cluster support. Operation of the cluster is simple and transparent to the user.

### Interaction

KNIME Cluster Execution belongs to the KNIME.com family of products (see box below). In combination with the KNIME Server, for example, the value of the overall solution can be increased by adding remote workflow storage, user access classes, and web access functionality.

Contact KNIME.com at the address below to obtain more details about KNIME, KNIME Cluster Execution or the other products in the KNIME.com product family.

### Software & Licensing

**Submit Clients:**  
Linux SUSE 10/11, Fedora 10 Red Hat Ent. Linux 5 (32 and 64 bit)  
**Cluster Engine:**  
Sun Grid Engine (SGE 6.2)  
**Cluster Slaves:**  
Linux SUSE 10/11, Fedora 10 Red Hat Ent. Linux 5 (32 and 64 bit)

**Licensing**  
KNIME Cluster Execution is available under an annual license model based on the number of users. It includes professional support for KNIME Desktop.

Cluster Execution

KNIME Server

KNIME Team Space

KNIME Professional

KNIME Desktop

### The KNIME Product Suite

- KNIME Desktop**  
Open-source platform for integrated data access, data mining, statistics, visualization, and reporting.
- KNIME Professional**  
Professional support on all KNIME features and priority bug fixes through a commercial agreement.
- KNIME Team Space**  
All KNIME Professional features plus shared workflow store, shared data space, shared metanodes.
- KNIME Server**  
All KNIME Team Space features plus remote and scheduled execution, user access rights, SOA integration, configurable web portal access to workflows and reports through a web browser.
- KNIME Cluster Execution**  
Submission of workflows to a compute cluster, distribution of single nodes, splitting of large data sets, and remote execution.

#### About KNIME and KNIME.com AG

KNIME is the leading open-source analytics platform and is based on a modern graphical workflow paradigm. From day one, rigorous professional software engineering processes combined with cutting-edge data analytics and visualization techniques have been applied to produce state-of-the-art results. KNIME is used by professionals in both industry and academia in over 60 countries. Open-source KNIME, as well as the KNIME product suite are supported and continue to be developed by KNIME.com AG, an independent vendor located in Zurich, Switzerland. See [www.KNIME.com](http://www.KNIME.com) for more details.



**KNIME.com AG**  
Technoparkstr. 1  
8005 Zurich  
Switzerland

**info@KNIME.com**  
**http://www.KNIME.com**  
**Tel: +41 44 445 2660**  
**Fax: +41 44 445 2662**