From the desktop to the grid and cloud: conversion of KNIME workflows to WS-PGRADE

Luis de la Garza Corresp. 1, Fabian Aicheler 1, Oliver Kohlbacher 1, 2, 3, 4
1 Center for Bioinformatics, Department of Computer Science, University of Tübingen, Tübingen, Germany
2 Faculty of Medicine, University of Tübingen, Tübingen, Germany
3 Quantitative Biology Center, University of Tübingen, Tübingen, Germany
4 Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen, Germany
Corresponding Author: Luis de la Garza
Email address: delagarza@informatik.uni-tuebingen.de

Computational analyses for research usually consist of a complicated orchestration of data flows, software libraries, visualization, selection of adequate parameters, etc. Structuring these complex activities into a collaboration of simple, reproducible and well defined tasks brings down complexity and increases reproducibility. This is the basic notion of workflows.

Workflow engines allow users to create and execute workflows, each having unique features. In some cases, certain features offered by platforms are royalty-based, hindering use in the scientific community.

We present our efforts to convert whole workflows created in the Konstanz Information Miner Analytics Platform to the Web Services Parallel Grid Runtime and Developer Environment. We see the former as a great workflow editor due to its considerable user base and user-friendly graphical interface. We deem the latter as a great backend engine able to interact with most major distributed computing interfaces. We introduce work that provides a platform-independent tool representation, thus assisting in the conversion of whole workflows. We also present the challenges inherent to workflow conversion across systems, as well as the ones posed by the conversion between the chosen workflow engines, along with our proposed solution to overcome these challenges.

The combined features of these two platforms (i.e., intuitive workflow design on a desktop computer and execution of workflows on distributed high performance computing interfaces) greatly benefit researchers and minimize time spent in technical chores not directly related to their area of research.