

Optimizing a Grid workflow for the EGEE infrastructure: The case of Wien2k

Wednesday, February 13, 2008 2:00 PM (20 minutes)

In real production grids the time between submitting a grid activity and its execution ranges from 10 to 60 minutes. When porting a complex workflow to the Grid, such as Wien2K, this overhead does not only appear once, but repeatedly, and increases the execution time of workflows largely.

In our previous presentation we showed aggregation of grid activities. Since then we have also experimented with other means of reducing overhead through scheduling:

Worker nodes may be scheduled without an active task, which they request through a pull model from a coordinator. These workers would be submitted once, and thus the overhead of scheduling them would only appear once. However, this mechanism is unfair towards other users, as it occupies resources for a longer period than one task. We will show results of our experiments, and classify them according to speed and fairness.

3. Impact

While similar research has been done for simple parameter studies, complex workflows have not yet been studied at this level of detail.

Although the optimization techniques shown here are applied to the Wien2K workflow in particular, they are generic enough to be applied to other complex Grid workflows. The lessons learned from porting the Wien2k application can provide a guidance for other future work: Porting other applications should become much easier, as the same patterns can be applied to other work.

The Wien2K application is currently used by thousands of scientists: Optimizing the Wien2K workflow for the Grid would enable these scientists to use the Grid for their calculations, resulting in significantly lower simulation time.

URL for further information:

<http://www.dps.uibk.ac.at/>

4. Conclusions / Future plans

We plan to continue to optimize and improve our Grid version of Wien2K. During the EGEE III project we plan to work towards the user: Providing a user-friendly interface through the use of a web portal, and providing an "easy" downloadable package, without violating the original Wien2k license. We are also continuously working on new ideas of workflow improvement.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Workflows, Computational Chemistry, Wien2K, Application Porting

1. Short overview

Wien2K is a package for electronic structure calculation of crystals. After porting Wien2K to the gLite environment we discovered a severe performance drawback: Scheduling in the real life EGEE infrastructure provides proved to take too long to manage a small workflows efficiently. We have already presented some optimization work through aggregation and will now present other optimization techniques.

Primary author: Dr BERGER, Maximilian (University of Innsbruck)

Co-author: Prof. FAHRINGER, Thomas (University of Innsbruck)

Presenter: Dr BERGER, Maximilian (University of Innsbruck)

Session Classification: Workflow and Parallelism

Track Classification: Application Porting and Deployment