Geosci. Model Dev. Discuss., 8, C876–C880, 2015
www.geosci-model-dev-discuss.net/8/C876/2015/

Geoscientific Model Development Discussions
Open Access

GMDD
8, C876–C880, 2015

Interactive Comment

Interactive comment on "Experiences with distributed computing for meteorological applications: Grid computing and Cloud computing" *by* F. Schüller et al.

F. Schüller et al.

felix.schueller@uibk.ac.at

Received and published: 27 May 2015

Reviewer comments are shown in bold. Please find the latexdiff version of the paper as supplement pdf.

The topic of this paper is very interesting, but, from my point of view, this manuscript do not address some of the most relevant issues found when running meteorological applications in distributed infrastructures. Furthermore, the results shown do not help to quantify the advantages of using each computing

infrastructure because each application is evaluated in a different infrastructure.

Unfortunately we do not have the resources to port our applications to both infrastructures. Quantifying the advantages therefor is not possible for us, as each of our applications has different needs. This is why we chose the word "Experiences" and not "Comparison" in the title. To make this clear right from the beginning, we included this information at the end of the introduction.

The paper has a good introduction. The description of the Grid and Cloud computing and the description of the Middleware ASKALON are also very good. I find sections Advantages/disadvantages Grid/Cloud difficult to read. From my point of view, it lacks of structure and the information is scattered. I suggest to analyze the same topics for both infrastructures and create a table brief.

We followed your helpful suggestion and added another table with only the main points of these 2 sections (now table 1) combined to give a brief overview.

Some important aspects of running meteorological application on Grid are missing in this section. I think this hap- pens because the Grid infrastructure used to perform this study seems to be reliable and homogeneous, and the Grid sites connected through a fast network. These are not common characteristics of Grid infrastructures. It is no rare to find Grid infrastructures connected through slow Internet connections, with different operating systems and where

GMDD

8, C876–C880, 2015

jobs sometimes crash unexpectedly.

Yes, the Grid we used consisted mostly of reliable and homogeneous hardware, but it also included ad-hoc networks build by clustering (students-) Desktop PC's during the night. We included this information at the description of the AGRid project (first paragraph section 2.1). We also argue that this characteristic is not uncommon: for example the grid initiatives EGEE, french Grid5K, ESGF, US - Teragrid have or had a similiar setup and network connection as AustrianGrid.

In these scenarios, special middleware to manage the monitoring of the simulations, the heterogeneity of libraries (including MPI) and the data management are required. In my opinion, this aspects should be mentioned in a paper that talks about Experiences with distributed computing for meteorological applications.

To overcome problems resulting from unexpected job crashes or slow network connections, our Middleware Askalon is able to handle these. We extended Askalon's description and added another reference where these capabilities are described in more detail.

ASKALON is able to handle most of the common failures as jobs and file transfers are resubmitted on failure and jobs might also be rescheduled to a different resource if transfers or jobs failed more then 5 times on a resource (Plankensteiner 2009a). These features still exist in the Cloud version but play a less important role as resources

GMDD

8, C876–C880, 2015

showed to be more reliable in the Cloud case.

The title of the reference is "A new fault tolerance heuristic for scientific workflows in highly distributed environments based on resubmission impact".

Regarding the description of the applications and the experiments: I do not understand what authors mean with the sentence: The workflow characteristics relevant for distributed computing are: fewer model instances but highly CPU intensive as well as lots of interprocess communications. Fewer compared to what?

What we meant is: fewer as compared to the other two applications. As this is unclear we changed the wording to : "few (2050)", reflecting the information from Table 2. "Overview of our projects and their workflow characteristics."

Regarding the sentence: Based on the experience from the MeteoAG experiments, hypothesize that it would be much more effective to deploy an application consisting of serial CPU jobs. Why would it be most effective to deploy an application consisting on serial CPU jobs?. I understand a simpler model would be faster, but why serial?

ASKALON is optimized for submission of single core parts of a workflow, which avoids internal parallelism of activities but allows best control over the execution within

GMDD

8, C876–C880, 2015

ASKALON. This is making it easier to use systems to capacity and also making it easier to move activities to different systems. So with Askalon it is more efficient for the whole deployment to have many serial activities than one internally parallelized activity.

In general, I find sections 3 and 4 difficult to read. The discussion about costs is too complex and I do not think it is very useful for the reader.

We completely reworked the section about costs, hopefully making it easer to read.

Mistakes: -In the line 9 of section 3.3 RainCloud, the word be is twice. -In the caption of Figure 1 right and left are changed.

Thank you for catching these mistakes, we corrected them.

Please also note the supplement to this comment:
http://www.geosci-model-dev-discuss.net/8/C876/2015/gmdd-8-C876-2015-supplement.pdf

Interactive comment on Geosci. Model Dev. Discuss., 8, 1171, 2015.