Geoscientific
Model Development
Discussions

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

Anonymous Referee #1

Received and published: 16 April 2015

This paper shows experiences with three meteorological applications in Grid and Cloud infrastructures. The authors describe how these applications were ran in distributed infrastructures and conclude that distributed infrastructures can complement and extend high performance computing.

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.

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. Some important aspects of running meteorological application on Grid are missing in this section. I think this happens 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 jobs sometimes crash unexpectedly. 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.

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?

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?

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.

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.