Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-193-AC2, 2016

Geoscientific Model Development Discussions

Interactive comment on "Enabling BOINC in Infrastructure as a Service Cloud Systems" *by* Diego Montes et al.

Diego Montes et al.

kabute@uvigo.es

Received and published: 23 December 2016

Thank you a lot for your review and comments. Answering your specific questions:

How the use of AWS compares against using other computational resources available (grid computing, supercomputers, clusters, domestic resources,...), in different aspects, including but not limited to the costs. For example the total time to get a full ensemble simulation including the pre and post-processing.

We have completed comparisons with the Oxford Supercomputer ARCUS-B (Uhe et al., Utilising Amazon Web Services to provide an on demand urgent computing facility for climateprediction.net, Proceedings of the 2016 IEEE 12th International Conference on e-Science), where it is shown that running the simulations on the supercomputer

is more than twice as expensive than AWS spot instances but half the price that on-demand. Run times are similar, and this can be extrapolated to other supercomputers (of course depending on variables such as size, utilization and queueing on the resources). For our work we understand that "cluster" is an abstract and high level concept that include (and can be constructed on the top of) solutions over cloud computing. Compared with domestic resources (volunteers), it typically takes 7 times more (than AWS) in order to get an 80% of results returned while AWS returns a 100%.

Compare the use of commercial cloud providers to other scientific cloud providers like EGI FedCloud

A comparative of different cloud providers, is part of the aforementioned work and will be published in further papers but we would like to note that one of the coauthors of the paper (David Wallom) has participated actively in the design of the EGI FedCloud. Therefore, we are well aware of the advantages and limitations that it could have. The EGI Federated cloud is, as its name suggests, a federation of heterogeneous resource providers. So, they have all made different choices as to the infrastructure that they provide and the cloud management stacks they use to provide the user facing clouds services. Therefore we would actually only be testing any single system within that federation and so it would not be neither fair nor accurate to say that this is a performance benchmark of the 'EGI federated cloud'. We do also note though that due to its nature we may in future (if there are enough services providers with common hardware systems underpinning different choices in cloud management stack be able to test and compare performance of these different private cloud software solutions.

Could this approach be useful not only for Ensemble Predictions?

Our approach can be used for ensemble predictions as discussed in related works, but the main goal of this paper is to show that the model can be ran over the cloud and that BOINC can be a valid resource manager for any model. Sure, it could be used for

other problems in the field of climate research beyond ensemble studies

Include more details about the costs of moving and archiving the data. If every simulation generates 656 GB of data what is the costs of running many (ensemble) simulations? Include numbers

As mentioned, the cost of the data storage and transfer is non-existent for this work but have included specific numbers and prices in the final version of the paper (including S3 and Glacier). Also, it is worth to note that there is an AWS fee waiver for academic institutions that, usually, makes data transfer free.