Running Large and Dynamic Climate Ensembles in the Cloud

Diego Montes[1], Juan A. Añel[1], Javier Rodeiro[2]

email: kabute@uvigo.es

[1] University of Vigo, EPhysLab, Ourense, Spain, [2] University of Vigo, School of Informatics Engineering, Ourense, Spain

Environmental | Physics | Laboratory
EPhysLab
UniversidadeVigo

Problem Background

- Climate and meteorological oriented software and simulations usage in Cloud Systems is growing and has proven to be a feasible and solid solution for these kind of scientific applications.
- This approach allows scientists to reduce toil and burden (e.g. infrastructure setup and troubleshooting).
- Our most recent work has the focus on providing an out-of-the-box solution for WRF that works on diverse environments, specially public clouds.

Goals

- Provide an idempotent out-of-the-box solution that could be reused in a wide range of environments to run Climate simulations.
- Create a simplified workflow so, scientists using this solution can focus their efforts on doing science rather than setting up hardware or software that adds little value to their research.

Proposed Solution

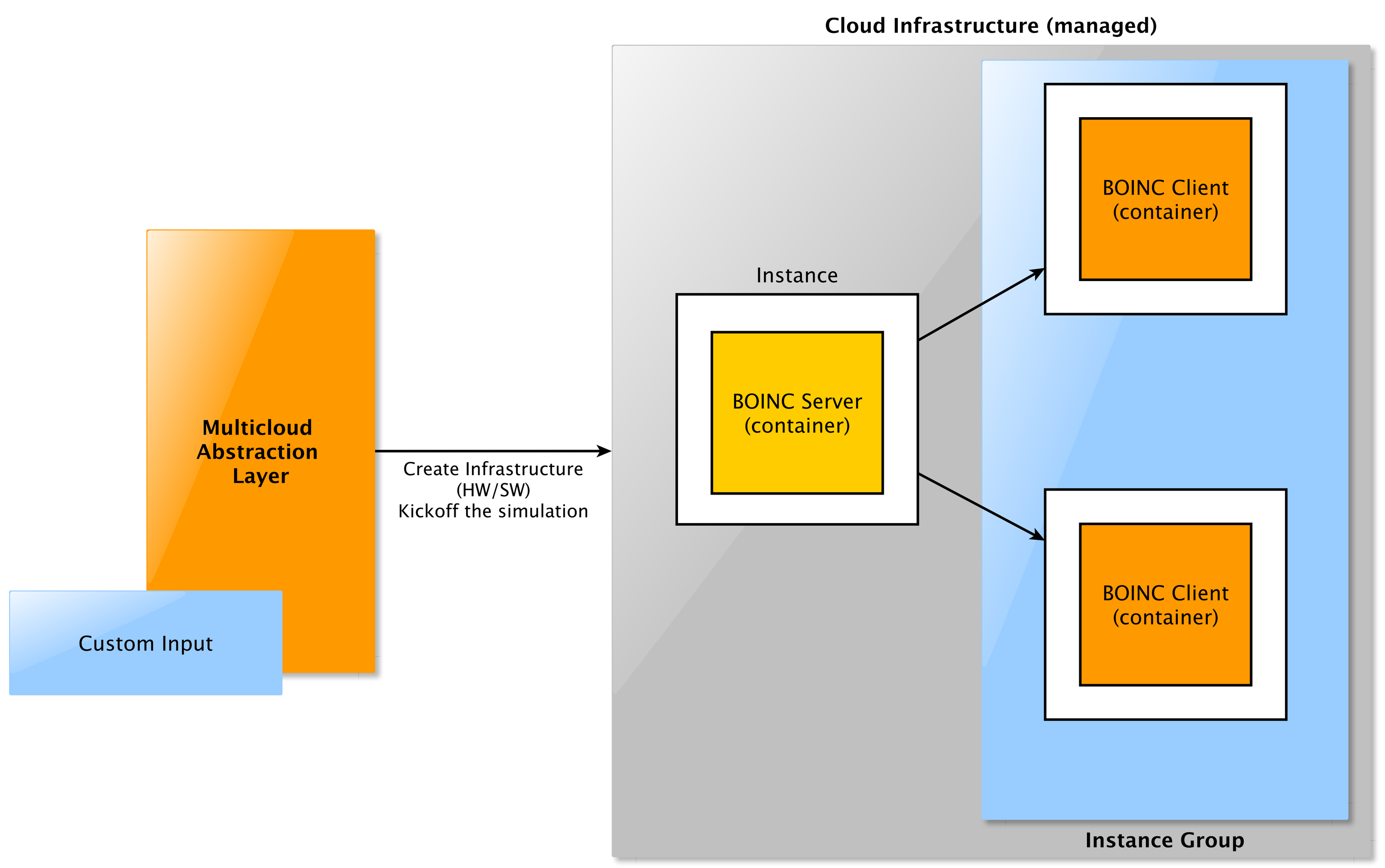

Figure 1. Proposed Solution Highe Level Architecture.

Our proposal is a highly abstracted multi-cloud solution with the following components:

- A **Multicloud abstraction layer** that uses in-house developed software that generates (from given custom inputs) the needed outputs (e.g. configured docker images and templates) so they can be consumed in an infrastructure as code solution and create the needed artifacts (both software and target infrastructure) and run the simulation.
- **Fully automated infrastructure** that runs WRF inside BOINC (by using the usual server/client architecture) with the inputs that the management layer created.

Results and Findings

- The multi-cloud management layer provided us the flexibility to advance on our previous work and embed our aforementioned BOINC solution on an environment that generates as an output a very well defined artifact: a fully automated, idempotent and immutable "Simulation as a Service" infrastructure to run climate ensembles.
- Resources definition and allocation (e.g. scalability) are done on setup time and via configuration variables.
- The usage of one container per instance reduced performance loss and provided scalability and simplicity to the solution.
- As we discussed on previous works the size of the simulations (we run one single simulation per container) needs to be adequate for the compute capacity of the underlying instance.
- Test and verification on Google Cloud Platform (GCP), running the ideal case (*em_tropical_cyclone*, 30 days simulated):

Instance Type	Specs	
n1-standard-1	16 vCPUs Xeon@2.30GHz (Haswell) and 60 GB of memory	32.50
n1-standard-16	16 vCPUs Xeon@2.30GHz (Haswell) and 60 GB of memory	4.01
n1-standard-64	64 vCPUs (simulation used 48 cores) Xeon@2.30GHz (Haswell) and 240 GB of memory	2.40

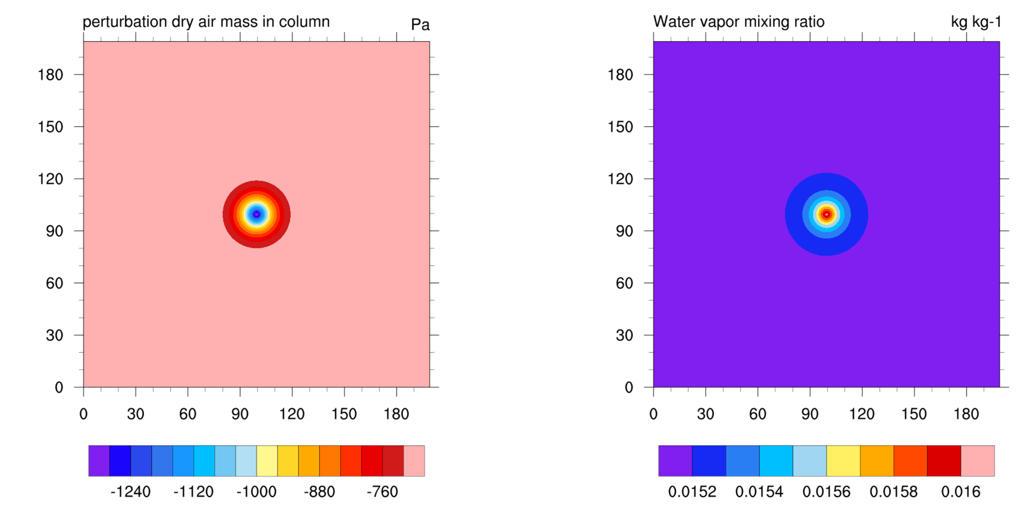

Figure 2. Ideal case used for architecture verification: *em_tropical_cyclone*, 30 days simulated.

References

[1] Montes, D.A. et al Enabling BOINC in infrastructure as a service cloud system *Geosci. Model Dev., 10, 811-826, https://doi.org/10.5194/gmd-10-811-2017, 2017.*

[2] Montes, D.A. et al Insights on the use of WRF in Cloud Systems *Geophysical Research Abstracts. Vol. 20, EGU2018-16881, 2018. EGU General Assembly 2018*

Acknowledgements

This material is based upon work supported by Google Cloud.