Views on the potential of Cloud computing and IaaS/HPCaaS for meteorology and climatology

Juan Antonio Añel, Diego A. Perez Montes, David C.H. Wallom, Peter Uhe, Pablo V. Caderno, Alberto Arribas, and Tomás F. Pena

Affiliations

Juan Antonio Añel and Diego A. Perez Montes: **EPhysLab, Universidade de Vigo**, Ourense, Spain (j.anhel@uvigo.es).

David C.H. Wallom: **OeRC, University of Oxford,** OX1 3QG, Oxford, UK,

Peter Uhe: **Environmental Change institute, University of Oxford**, OX1 3QY, Oxford, UK.

Pablo V. Caderno:**mAdme Technologies Ltd,** 32 Lower Lesson Street, Dublin 2, Dublin, Ireland,

Alberto Arribas: **Met Office Informatics Lab**, Exeter, UK

Tomás F. Pena: **CITIUS, University of Santiago de Compostela**, Santiago de Compostela, 15782, Spain

Goals

- Show areas of convergence between Voluntary Computing, Cloud and HPC.
- A practical example (part of our work).
- Comments on security, costs and financing.

Content

- Concepts and descriptions.
- BOINC overview.
- Examples: CPDN on-demand and WACCM over GCE.
- Security and Cloud.
- Cloud and financing.

Volunteer Computing

Problem solving oriented computing (e.g. simulations) or operations using virtual resources voluntarily donated by individuals or organizations.

- Existence of a middleware that abstracts the shared resources.
- Massively distributed.
- Redundant (one job is done N times, at least 1).

Volunteer Computing

Pros	Cons
High level of abstraction of the resources (highly independent of the hardware), thanks to the middleware.	Heterogeneous resource management.
Reduction of Costs (revert to the volunteer).	Unknown simulation/computing time (although there is usually a maximum time).
	In general, that is "best effort".

Cloud

A broad concept that encompasses the abstraction of computer systems (including their resources) and presents them in a ubiquitous way.

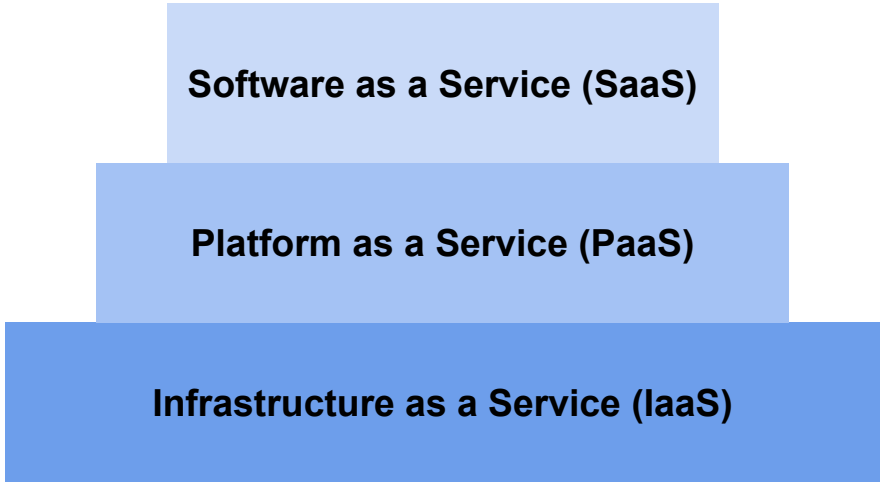
Public	Private
amazon web services™ Microsoft Azure Google Cloud vmware®	hp CESGA

Cloud

Pros	Cons
High abstraction ("everything is an API").	“Vendor lock-in” by using higher abstractions.
Fast scalability (“commodity hardware”).	It requires changing the implementation paradigm (including development).
In many cases cost reduction.	Incorrect use can lead to high costs.

Cloud

Types of Service (basic, from highest to lowest level of abstraction):



HPC (Traditional)

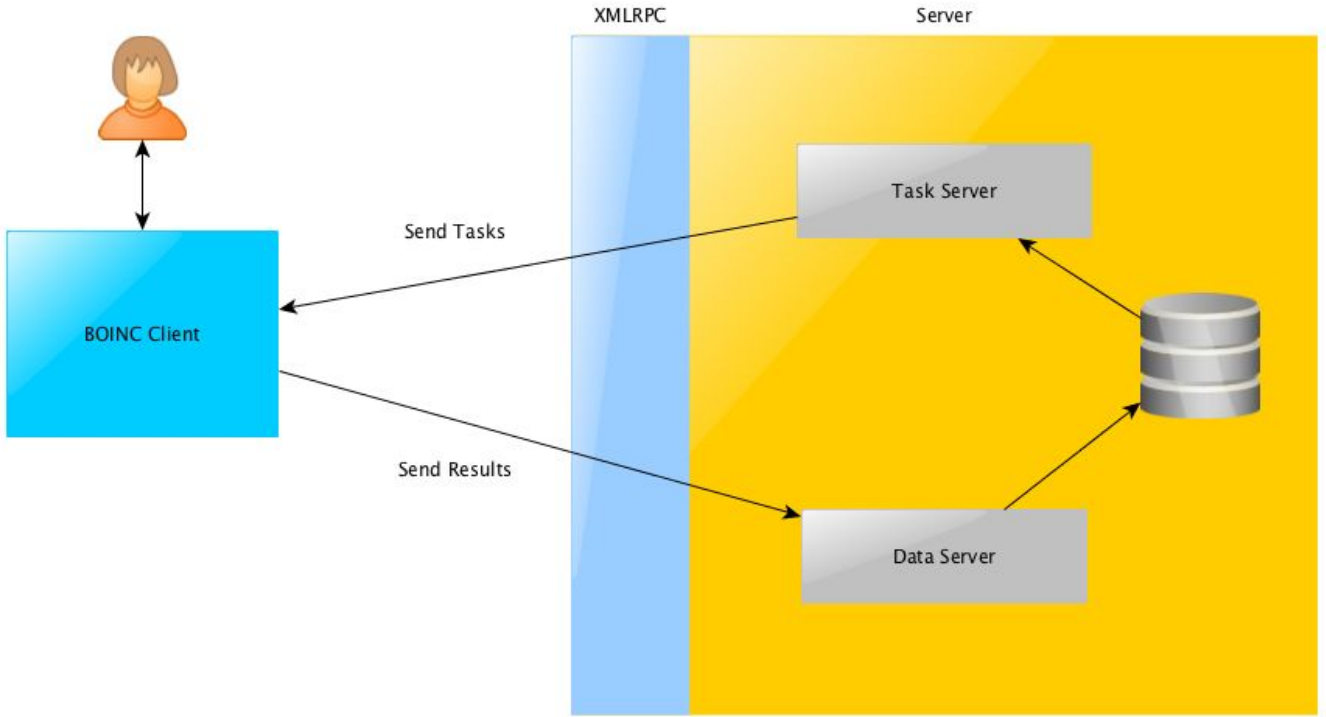
In-house: computing or supercomputing centers.

- Cost normally assumed by public entities.
- Better knowledge and proximity to the infrastructure.

BOINC

Berkeley Open Infrastructure for Network Computing. Probably the most well-known voluntary computing middleware. Base for projects such as: SETI @ home, LHC @ home, Climateprediction.net ...

BOINC


XMLRPC
Server
Task Server
Send Tasks
BOINC Client
Send Results
Data Server

CPDN on-demand

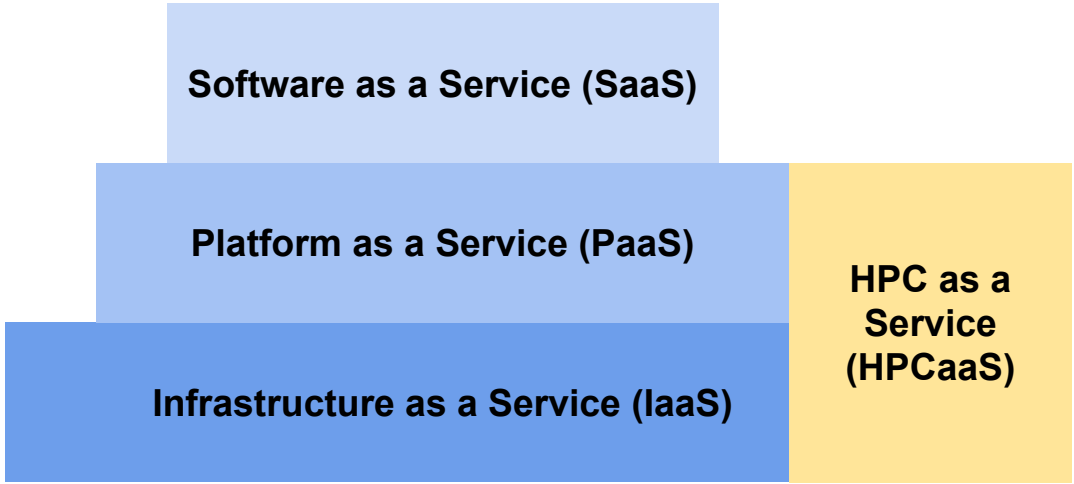
Climateprediction.net (CPDN) is the largest climate modeling experiment in the world, managed by Oxford University. The project runs thousands of simulations using voluntary computation over BOINC.

CPDN on-demand

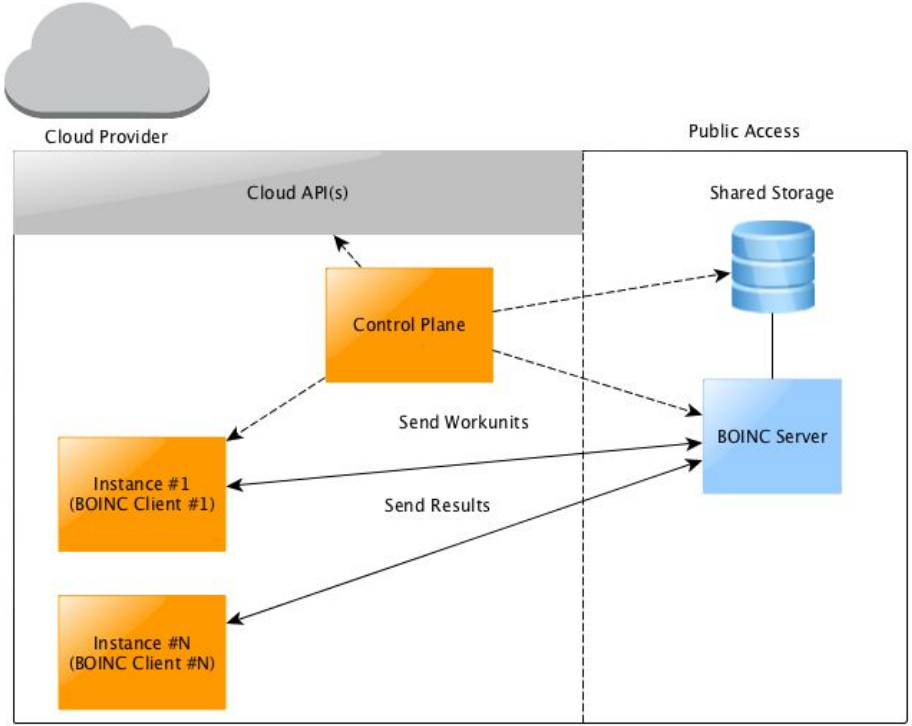
The goal was to provide to the CPDN a dynamic and reliable infrastructure, able to run massive simulations in a smaller and controlled amount of time.

For this we chose to migrate the infrastructure to the cloud (in our case AWS and Azure) and go towards HPC as a Service (HPCaaS).

CPDN on-demand



CPDN on-demand

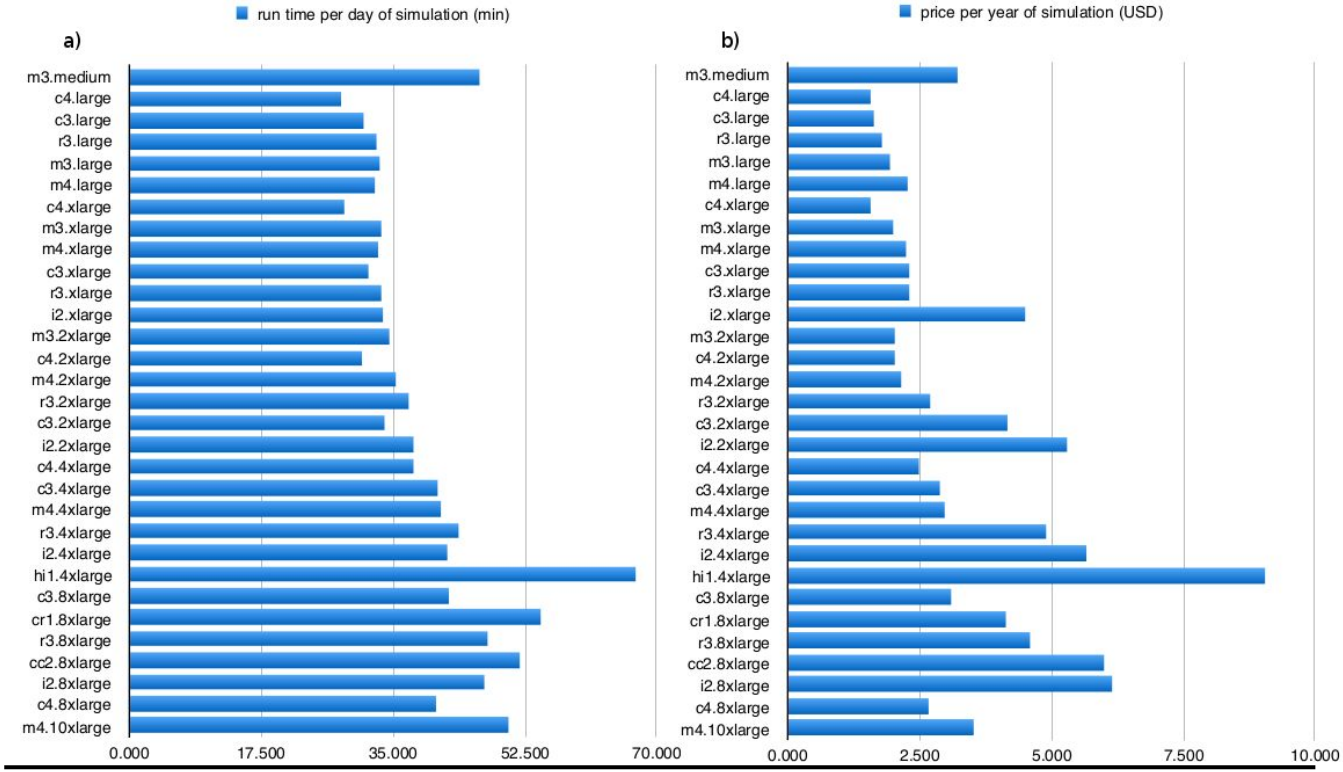


CPDN on-demand

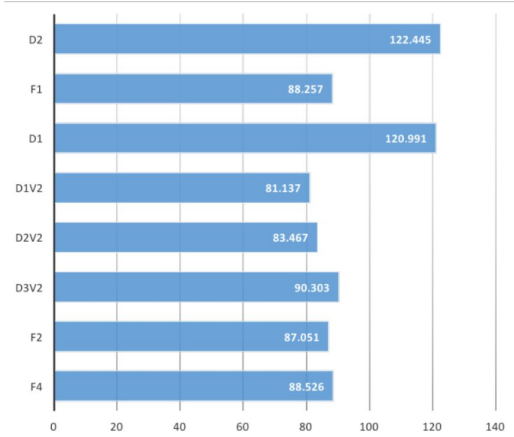
Pros:

- Using Cloud allows you to know the type of hardware and make forecasts (homogeneous environment).
- Use BOINC directly on 100% of resources.
- We mitigate the disadvantages of voluntary computing.

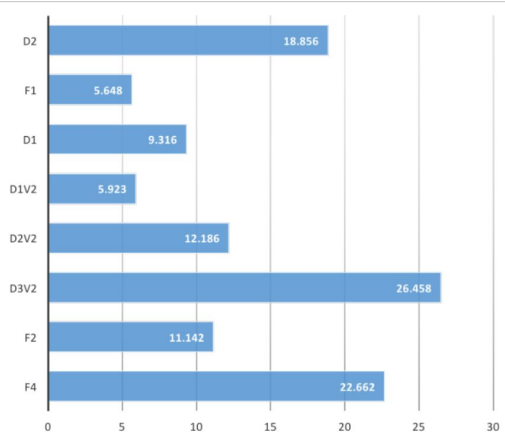
Some Data


run time per day of simulation (min)
a)
m3.medium
c4.large
c3.large
r3.large
m3.large
m4.large
c4.xlarge
m3.xlarge
m4.xlarge
c3.xlarge
r3.xlarge
i2.xlarge
m3.2xlarge
c4.2xlarge
m4.2xlarge
r3.2xlarge
c3.2xlarge
i2.2xlarge
c4.4xlarge
c3.4xlarge
m4.4xlarge
r3.4xlarge
i2.4xlarge
hi1.4xlarge
c3.8xlarge
cr1.8xlarge
r3.8xlarge
cc2.8xlarge
i2.8xlarge
c4.8xlarge
m4.10xlarge
0.000
17.500
35.000
52.500
70.000
price per year of simulation (USD)
b)
m3.medium
c4.large
c3.large
r3.large
m3.large
m4.large
c4.xlarge
m3.xlarge
m4.xlarge
c3.xlarge
r3.xlarge
i2.xlarge
m3.2xlarge
c4.2xlarge
m4.2xlarge
r3.2xlarge
c3.2xlarge
i2.2xlarge
c4.4xlarge
c3.4xlarge
m4.4xlarge
r3.4xlarge
i2.4xlarge
hi1.4xlarge
c3.8xlarge
cr1.8xlarge
r3.8xlarge
cc2.8xlarge
i2.8xlarge
c4.8xlarge
m4.10xlarge
0.000
2.500
5.000
7.500
10.000

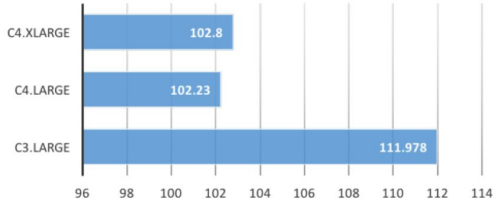
Some Data



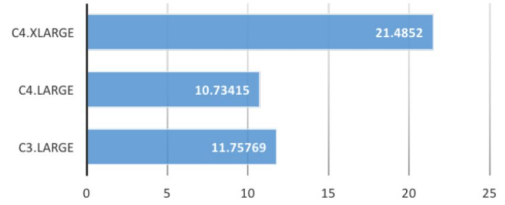
(a) Azure: Time per Workunit (hours)



(b) Azure: Price per Workunit (USD)

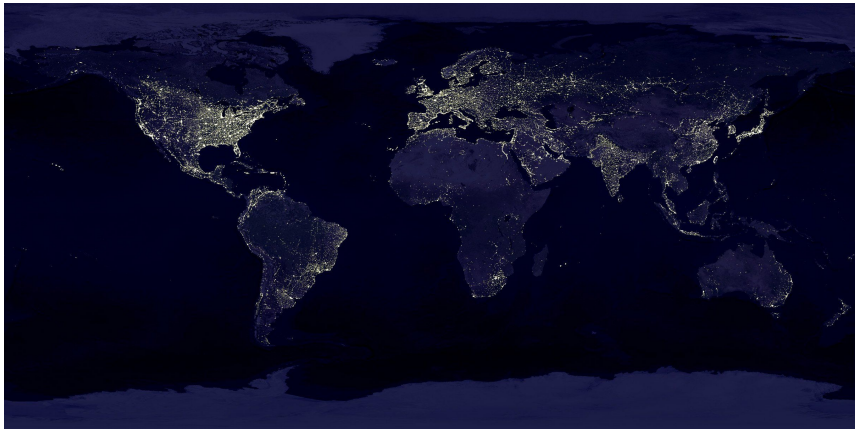


(c) AWS: Time per Workunit (hours)



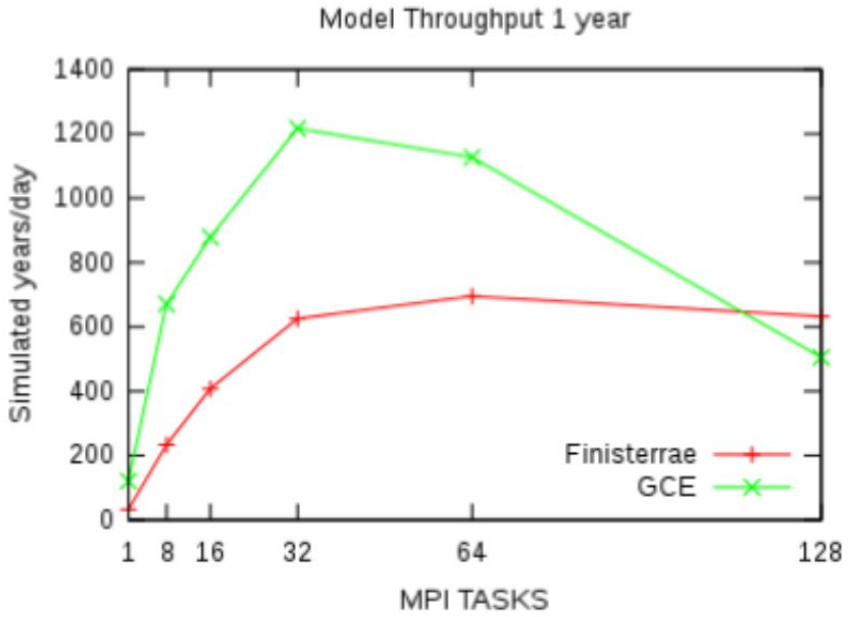
(d) AWS: Price per Workunit (USD)

WACCM over Google Cloud



The goal of this experiment (Caderno P. et al, 2016) was to migrate the Whole Atmosphere Community Climate Model (WACCM) to Google Cloud and compare its performance with the version installed in the Finisterrae Supercomputer.

WACCM over Google Cloud



Conclusions about the experiments in the Cloud

- Invested time to get an infrastructure running (with the models) drops dramatically.
- The reproducibility of the experiments is easier.
- We were able to "play" with a bigger variety of hardware and configurations.

Security in the Cloud

- There is a (false) perception that the Cloud (and data stored there) is not secure.
- Most providers are compliant with all regulations and certifications (eg ISO27001), not always available in other solutions.
- The integrity of the data (input and output) is especially important.
- The cloud is safe for scientific reproducibility (if used correctly).

Cloud and Costs

- Computational resources in the Cloud, although they may be more economical (if correctly used) remain expensive for some researchers.
- Special features are available (e.g., low latency networks) but this increases costs.
- It is possible (and advisable) to include Cloud in the "mix" if variable (and scalable) resources are needed.
- Growth in the funding of scientific projects in the Cloud is expected. For example EU Horizon 2020

Takeaways

- Both Voluntary Computing and Cloud require a change in the paradigm and the way of working (e.g. writing the code).
- There are multiple benefits to using a solution that mixes the three proposals.
- The costs of the Cloud are similar (or less) to the traditional HPC solutions, but they propose a different financing model.
- The Cloud is secure and provides a very good opportunity for the scientific community.

Questions?

References

- Montes D. et al., 2017: Enabling BOINC in infrastructure as a service cloud system, Geoscientific Model Development, 10, 811–826,. (doi:10.5194/gmd-10-811-2017)

- Montes D. et al., 2017: Use of several Cloud Computing approaches for climate modelling: performance, costs and opportunities. Proceedings of the EGU2017.

- Vazquez-Caderno, P., 2015: Usage of Google Compute Engine for climate studies with global climate models. M.S. thesis, Universidade de Santiago de Compostela.