Running GEOS-Chem on Cloud Computing Platforms

Jiawei Zhuang[1], Robert M. Yantosca[1], Sebastian D. Eastham[1], Daniel J. Jacob[1]

[1]Harvard University, Cambridge, MA, United States

1. Abstract

We have proven that GEOS-Chem is able to run on cloud computing platforms such as Amazon Elastic Compute Cloud (EC2). This will allow users to run long-term GEOS-Chem simulations and store large amounts of output data without having to invest in local computing infrastructures. Users will also be able to launch a virtual system with all compilers and libraries correctly installed, without worrying about software configurations. A basic tutorial is available online to help users to perform a short GEOS-Chem simulation on the cloud.

2. Cloud computing in atmospheric science

Cloud computing provides tremendous computational resources and storage to support model simulations and data processing. The price of cloud resources can be attractive, especially when one cannot ensure stable utilization of local machines[1].

The Weather Research and Forecasting Model (WRF) and the Community Earth System Model (CESM) have recently been made able to operate on cloud computing platforms[2][3]. Cloud platforms also facilitate data storage and sharing. NASA-NEX[4] is one of the public data sets held by Amazon Web Service (AWS) that provide high-resolution climate and weather data.

3. Make GEOS-Chem compatible with free software

GEOS-Chem v11-01 or newer is compatible with the free, open-source GNU Fortran compiler[5] (Table 1). This update allows GEOS-Chem to run on any computational platforms, including various operating systems in cloud.

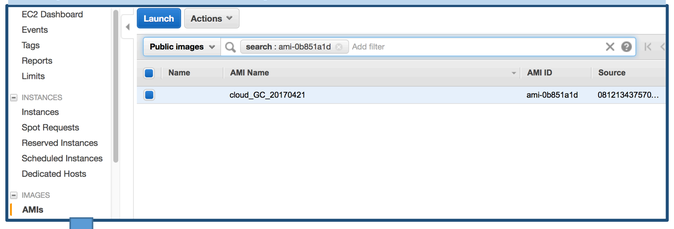
	Intel Fortran Compiler	GNU Fortran Compiler
Version	ifort 11.1.069	gfortran 6.2.0
Mean OH concentration (10^5 molec/cm^3)	12.312810	12.312863
MCF lifetime w.r.t. tropospheric OH (years)	5.1159	5.1159
Number of CPUs	24	
CPU type	Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50 GHz	
Memory used	5.2038 GB	4.5237 GB
Wall Time	04:22:54	05:07:26

Table 1. 1-month benchmark simulation for GEOS-Chem v11-02a. GNU Fortran gives the same simulation results as Intel Fortran, although with slightly lower performance.

4. Proof of concept -- a tutorial for everyone to run GEOS-Chem on Amazon Elastic Compute Cloud

GEOS-Chem has recently been tested successfully on Amazon Elastic Compute Cloud (EC2)[6]. Users can follow a tutorial online to launch a cloud server and run a short GEOS-Chem simulation[7]. Figure 1 demonstrates the basic workflow of using cloud platforms to perform GEOS-Chem simulations.

Step 1: In Amazon-EC2 console, launch a virtual server in cloud with pre-configured environment that allows GEOS-Chem to compile and run smoothly.



Step 2: Log in to that server with the tranditional "ssh" method, and perform the simulation just like on any other Linux servers.

```
$ssh $chmod 400 test_GEOSChem.pem
$ssh $ssh -Y -i "test_GEOSChem.pem" ubuntu@ec2-34-201-27-120.compute-1.amazonaws.com
The authenticity of host 'ec2-34-201-27-120.compute-1.amazonaws.com (34.201.27.120)' can't be established.
ECDSA key fingerprint is SHA256:ASQBaTZL4ppd0lKl5q7U5cv040kQ7IjJnuodfa/+sCM.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-34-201-27-120.compute-1.amazonaws.com,34.201.27.120' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-1013-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

9 packages can be updated.
0 updates are security updates.


Last login: Fri Apr 21 21:38:08 2017 from 65.112.10.153
ubuntu@ip-172-31-8-119:~$ cd GCv11-02/
compile.sh  gc_bleeding_edge  geosfp_4x5_standard  ut_bleeding_edge
ubuntu@ip-172-31-8-119:~/GCv11-02$
```

Step 3: Terminate that server when you finish the work. You are only charged when you are running the model.

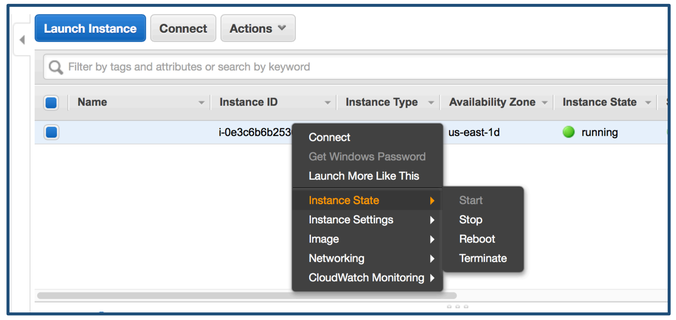


Figure 1. Main steps for running GEOS-Chem on the cloud

5. Compare with traditional platforms

Local machines	Cloud computing platforms
Need initial investment of computing infrastructures	Users can launch or delete a virtual server anytime
Need to build libraries and configure system environments	Users can get a copy of an existing system with everything installed
A machine has fixed capacity, which can become either redundant or insufficient over time	Users can dynamically control the number of CPUs and the amount of storage
Some local machines cannot access remote data efficiently	Data can be shared easily through the cloud
High bandwidth network between compute nodes for massively parallel MPI programs	Need extra effort and careful network tuning for MPI programs on a large number of cores

Table 2. Comparison between traditional machines and cloud platforms. Advantages are shown in green while disadvantages are in red.

Cloud computing platforms provide much greater flexibility in controlling computational resources (Table 2). However, while a program can easily parallelize over up to 64 cores in a single compute node, it takes much more effort to run Message Passing Interface (MPI) programs on thousands of cores across multiple nodes in cloud. This could be a challenge for the MPI-capable, High-Performance Version of GEOS-Chem (GCHP).

6. Conclusions and future directions

We have demonstrated the capability of GEOS-Chem to run on the Amazon EC2 cloud computing platform. However, a more concrete strategy is needed for long-term simulations and efficient data sharing in cloud. We also hope to cooperate with other community models like CESM and WRF to build a system in cloud where people can use all these models directly.

Acknowledgments. This work was funded by NASA ACMAP.

7. Reference

[1] Freniere, C., et al., *The Feasibility of Amazon's Cloud Computing Platform for Parallel, GPU-Accelerated, Multiphase-Flow Simulations.* Computing in Science & Engineering, 2016. **18**(5): p. 68-77.

[2] Molthan, A.L., et al., *Clouds in the cloud: weather forecasts and applications within cloud computing environments.* Bulletin of the American Meteorological Society, 2015. **96**(8): p. 1369-1379.

[3] Chen, X., et al., *Running climate model on a commercial cloud computing environment: A case study using Community Earth System Model (CESM) on Amazon AWS.* Computers & Geosciences, 2017. **98**: p. 21-25.

[4] https://aws.amazon.com/nasa/nex/

[5] http://wiki.seas.harvard.edu/geos-chem/index.php/GNU_Fortran_compiler

[6] https://aws.amazon.com/ec2/

[7] https://github.com/JiaweiZhuang/cloud_GC, can also be access by the QR code at the upper-right corner of this poster.