Practical Examples with Cloud Computing and OpenEye virtual screening tools

Hideyuki Sato PhD, OpenEye Japan

Abstract

Virtual screening has been implemented in a High Performance Computing (HPC) cloud platform. OpenEye tools run smoothly on the Cloud with StarCluster and OpenEye standard license. All tested tools; OMEGA, ROCS and OEDocking, work correctly. Various databases were used for these cases; e.g. DUDE, Namiki, eMolecules. For these test cases, Amazon Web Services (AWS) was used as the cloud computing platform. Cloud computing dramatically reduces the time for setting up a cluster and the total costs for computer resources.

Introduction

High Performance Computing (HPC) is one of the key technologies in modern drug discovery, because we have to access huge number of molecules and targets. To access large computer resources, computer clusters were introduced into each research site. However, having and maintaining such large clusters requires large budgets for hardware and maintenance costs, although the clusters are not fully used. In particular, because data base size for virtual screening is quickly growing, it is getting more difficult for pharmaceutical companies to design appropriate hardware in-house. Instead, HPC by cloud computing is becoming popular in drug discovery research, even within the pharmaceutical industry, because of its flexibility. OpenEye license policy has advantages for such up-to-date needs in drug discovery because it is flexible and offers unlimited usage. Therefore, we recommend all OpenEye users to use cloud computing for our virtual screening tools.

As cloud computing system, Amazon Web Services (AWS) was used. In order to configure a cluster on AWS, StarCluster[1] was used.

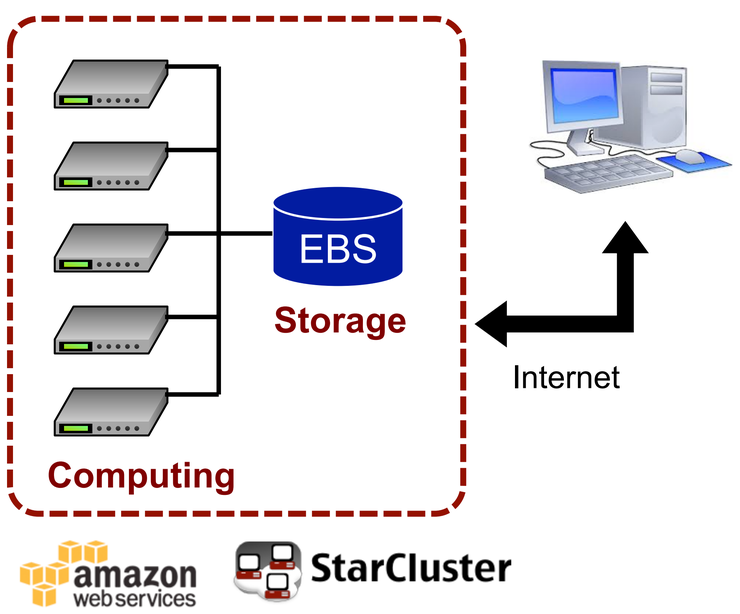

Fig. 1. HPC cluster on Cloud computing using Amazon Web Services (AWS) and StarCluster

StarCluster: Open source Python library for building cluster on AWS
Storage: AWS Elastic Block Store (EBS)
Computing: AWS Elastic Compute Cloud (EC2)

Method

A StarCluster image, ami-52a0c53b, which is based on Ubuntu 12.04, was used.
Various computing resources (instances) are available on AWS EC2. Instances were chosen depending on the specific purpose.

Table 1. Examples of instances which are available in AWS EC2

Instance	vCPU	GPU	Cost (per hour)
c3.large	2	-	$ 0.105
c3.4xlarge	16	-	$ 0.840
c3.8xlarge	32	-	$ 1.680
g2.2xlarge	8	1	$ 0.650
g2.8xlarge	32	4	$ 2.600

Programs

OMEGA 2.5.1.4, ROCS/subROCS 3.2.0.4, FastROCS 1.4.0.11, FRED 3.1.0.a and HYBRID 3.1.0.a were used.

Database preparation

The Namiki 201410HTS database was used for the speed test of each tool. The database was converted using sdfrename.py, one of the OEChemTK examples, to get the molecule name from the NAMIKI_ID SD tag. The database was pre-filtered with FILTER using the default parameters (BlockBuster filter, -pkanorm true etc.).

The OMEGA calculation was run with –maxconfs 200 and 10. 4,564,174 molecules were input to the OMEGA calculation. Molecules which have unspecified stereochemistry were automatically removed in the calculation. As a result, OMEGA database has 3,212,617 molecules with 457,859,198 conformers.

This OMEGA calculation was completed with ten c3.8xlarge instance of AWS, which have 320 cores as a cluster, and took 69 min.

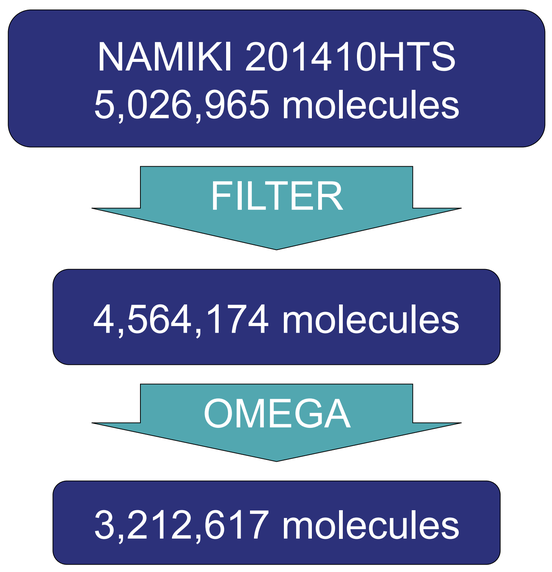

Fig. 2. A workflow to generate OMEGA database from Namiki 201410HTS

Speed test results

The cluster on AWS was created using StarCluster, and it consists of two c3.8xlarge EC2 instances which have 64 slots (cores). The database was saved on EBS.
For example runs the PDB structure 3eml (Adenosine A2a receptor) was used. Its ligand was used as a query for ROCS and subROCS, its protein was used for FRED and HYBRID calculations.
OpenMPI, included in each program, was used for all parallel calculations.

Table 2. Speed test results using virtual screening tools against Namiki 201410HTS

Program	Time	Instance	Cores
FRED	49 h 6 m	c3.8xlarge	64
HYBRID	23 h 57 m	c3.8xlarge	64
HYBRID	19 h 41 m	c3.8xlarge	320
ROCS	1 h 44 m	c3.8xlarge	64
subROCS	13 h 18 m	c3.8xlarge	320
FastROCS [1)]	load: 420 s search: 27 s	g2.8xlarge	32 (4 GPUs)

1) Used small database with maxconfs 10 in OMEGA.
30,990,324 confs, but the number of molecules is the same.

Discussion

HYBRID works faster than FRED as it uses the bound ligand in a receptor for pose selection. When 5 instances were used, HYBRID calculation was faster than 2 instances, but the speed increase was not proportional between core numbers and speed.
subROCS is slower than ROCS as it uses many more start points for aligning molecules. For subROCS, increasing instances helped to reduce total calculation time. Since the FastROCS server needs to load all molecules into RAM, a smaller size database (maxconfs 10) was used.

Conclusion

All OpenEye's virtual screening products work very well on Cloud computing by Amazon Web Services with our standard license. StarCluster provides appropriate configuration of the cloud cluster for OpenEye products with Ubuntu 12.04. All the tested products work well with OpenMPI in each OpenEye product for parallel calculations.
Cloud computing provides flexible hardware environment for virtual screening with OpenEye products. OpenEye license policy fit needs for such flexible computer resources in drug discovery.

Acknowledgments

Craig Bruce gave a lot of useful advices on setting up AWS and StarCluster. James Haigh helped to run calculations.

References

[1] StarCluster - http://star.mit.edu/cluster/docs/latest/index.html

OpenEye Japan K.K.

AIOS Toranomon 904
1-6-12 Nishishimbashi
Minato-ku, Tokyo
105-0003, Japan

Tel. 03-6206-1425
oe_japan@eyesopen.com
www.eyesopen.com