4-21-2017

Computational Chemistry in the Cloud

Nathan Vance

Computational Chemistry in the Cloud

Nathan R. Vance and Dr. William F. Polik
Department of Chemistry, Hope College, Holland, MI 49423

Introduction

The compute cloud abstracts computer hardware into virtual machines hosted by companies such as Google and Amazon.

Google Cloud Platform

amazon web services™

The Server in the Cloud (SITC) project was developed to adapt the compute cloud to computational chemistry. SITC is a set of scripts that installs a WebMO computational chemistry webserver to a cloud computer. A single computational chemistry webserver, however, is not powerful enough to service research users. The CLuster In the Cloud (CLIC) project is a job scheduler that takes advantage of the cloud environment by spawning new cloud instances as needed to complete jobs.

The Compute Cloud

Anyone who has performed a Google search, watched a video on Netflix, or stored a file on Dropbox has used the cloud. In each case, a service is provided. For Dropbox the service is storage, Netflix provides movies as a service, and Google provides advertising as a service.

The Compute Cloud provides computing infrastructure as a service. Advantages of using the compute cloud include:

- Flexible choice for computer architecture
- Dynamic scaling
- No hardware maintenance
- No up front costs
- Pay for what you use, not what you're capable of using

Since the cloud's pricing follows the utility model rather than ownership model, significant cost savings can be achieved in situations with highly variable use. Additionally, the monetary and expertise barriers of entry associated with owning computing hardware are eliminated.

Build your own Cluster (BYOC)

We wrote a guide to purchase, setup, and configure computer clusters: www.webmo.net/support/byoc-centos7.pdf

Cost: $680,000

Pros:
- Tried and true technology
- High performance

Cons:
- High initial cost
- Expensive to run
- Difficult to maintain
- Overcompensates compute power to satisfy peak demand

Server in the Cloud (SITC)

We implemented a cloud-based computational chemistry server: https://github.com/NathanRVance/sitc

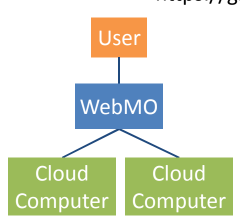
Cost: $25 / month

Pros:
- No hardware to maintain or upgrade
- Can dynamically select capabilities based on demand
- Zero initial cost

Cons:
- Pay for ongoing use

CLuster In the Cloud (CLIC)

We are implementing a cloud aware job scheduler: https://github.com/NathanRVance/clic



Cost: $0.035 / CPU hr

Pros:
- Dynamically expands and contracts grid to accommodate use
- Under high use, jobs are still processed quickly
- Pay only when in use

Cons:
- Economical when use is highly variable

Management of Dynamic Resources

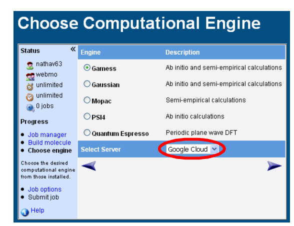
Since latency in creating new cloud instances exists, it is necessary to decide how to queue jobs and when to expand and contract the grid.

	Number Based	Time Based
Queueing	One queue for each virtual computer, new jobs go in shortest queue	One queue total, first job in line goes to next available server
Expansion	Expand when the queue is too long	Expand when jobs wait too long
Contraction	Contract the moment the queue empties	Contract after the queue remains empty for some time

Since computational chemistry jobs are highly variable in time, the number of jobs in the queue is meaningless in determining if the line is "long" or not. Therefore, CLIC focuses on the length of wait rather than the number of jobs in line.

Application: WebMO

We plan to integrate CLIC functionality into WebMO for on-demand dynamically scalable computing.



WebMO server selection

When a user submits a job, they can specify the desired capabilities of the server the job would run on. CLIC then creates a cloud instance matching these capabilities, the job is run on it, and the cloud instance is deleted.

Acknowledgements

- JR Schmidt
- WebMO LLC
- Beckman Foundation
- Hope College Chemistry Dept