

科学技術計算ワークフロー実行に向けた Amazon EC2 におけるインスタンス性能評価

小川光貴¹ 杉村太一² 小坂隆浩²

概要 : 近年, 商用クラウド環境の普及によりクラウド環境における科学技術計算ワークフローの実行が注目されている。本研究では Amazon Elastic Compute Cloud (AmazonEC2)より提供されているリソース(インスタンス)の取得から開放にかかる時間及び, リソースのコストパフォーマンスの 2 つの観点から適切なリソース選択を提案する。

キーワード : 科学技術ワークフロー, Amazon EC2, インスタンス

1. はじめに

科学分野において新たな科学的知識は, 科学者がデータ解析と知識発見のパイプラインを組み立てることによって得られることが多い[1]。クラウド環境での実行はリソースを共有している特性からリソースの要求を行い, 即座に開放される保証はない。本研究では AmazonEC2 を用いて 1 つのインスタンスで数分の間にリソースの取得と開放を繰り返し, 所要時間を評価することでクラウド環境での効率的な実行について確認する[2]。また, 科学技術計算ワークフローは複数のタスクから構成されており, 各タスクの実行時間, 入出力(I/O)などの特性が異なる。利用者にとって性能の差異によるリソース選択は問題点の 1 つである。先述した評価方法に加え, 各インスタンスの I/O とコストを比較することで, クラウド環境で科学技術計算ワークフローを実行するために適切なリソース選択を目的とする[3]。

2. クラウド環境実行についての評価

本章では 1 つのインスタンスで数分の間にリソースの取得と開放を繰り返し所要時間を評価する。

2.1 評価方法

本実験で取得する値としてインスタンスを起動し Instance State が実行中になるまでの時間(VM Deployment Time For Res. Acquisition(VDT))と Status Check が OK になるまでの時間(VM Boot Time For Res. Acquisition(VTT)), インスタンスを終了してから Instance State が停止になるまでの時間(Total Time For Res. Release(TT))を計測する。インスタンスタイプは t4g.small, t4g.large, t4g.xlarge, c6g.medium, c6g.xlarge とし, 5 種類のインスタンスタイプそれぞれに対し, 上記の時間を取得する実験を 20 回繰り返す。

2.2 評価結果

2.2.1 VM Deployment Time For Res. Acquisition

図 1 に各インスタンスタイプの VDT を示す。汎用型である t4g がコンピュータ最適化型である c6g よりも平均値が高い。t4g.small に着目するとばらつきが多い。

2.2.2 VM Boot Time For Res. Acquisition

図 2 に各インスタンスタイプ VTT を示す。全体的な平均として 150s 前後で Status Check が OK になっているため, インスタンスタイプによる Status Check の影響はない。

2.2.3 Total Time For Res. Release

図 3 に各インスタンスタイプの TT を示す。全体的な平均としては 50s 前後ではあるが, インスタンスサイズ xlarge に着目すると t4g では最小値がほかのものに比べて高いが c6g では最大値が他のインスタンスと比べて低い。

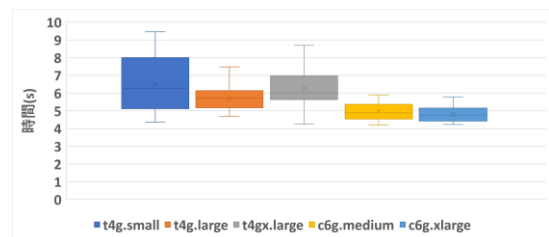


図 1 : 各インスタンスタイプの VDT

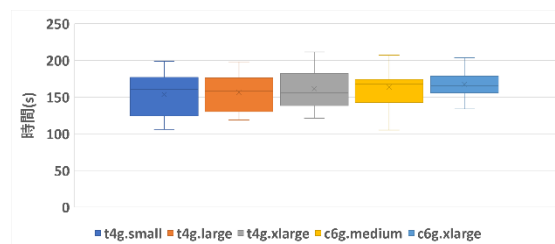


図 2 : 各インスタンスタイプの VTT

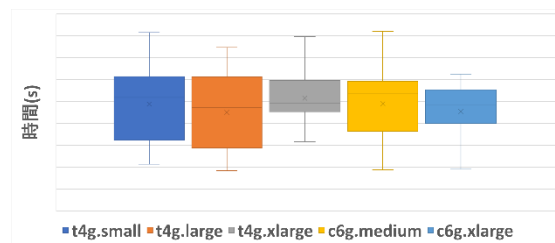


図 3 : 各インスタンスタイプの TT

¹ 同志社大学理工学部
Faculty of Engineering, Doshisha University

² 同志社大学大学院理工学研究科
Graduate School of Engineering, Doshisha University

3. 科学技術計算ワークフロー実行に向けたリソース選択

本章では異なる I/O 性能とコスト, 2 つの観点からコストパフォーマンスを評価する.

3.1 評価方法

ベンチマークソフト fio (Flexible IO Tester)を用いて, 設定ファイルを実行したときの単位時間あたりの入出力量 (IOPS) を Sequential-Read(SR), Sequential-Write(SW), Random-Read(RR), Random-Write(RW)の 4 項目と, AWS から提供されている各インスタンスのオンデマンド料金をそれぞれ比較する.

3.2 評価結果

3.2.1 Sequential-Read(SR)

図 4 に SR の比較結果を示す. コストの観点からみると t4g.large がファミリーサイズが小さいにもかかわらず IOPS の値がサイズが大きいものと変わらないため, コストパフォーマンスは高い.

3.2.2 Sequential-Write(SW)

図 5 に SW の比較結果を示す. SW においては c6g.medium に着目するとタイプは異なるが t4g.small よりもサイズは大きいにも関わらず IOPS の値が低い.

3.2.3 Random-Read(RR)

図 6 に RR の比較結果を示す. IOPS の値に着目するとインスタンスタイプ t4g が IOPS が高い. またコストの面で考慮すると t4g.small が中でも安価で高い IOPS の値を測定している.

3.2.4 Random-Write(RW)

図 7 に RW の比較結果を示す. SW 同様に c6g.medium に着目すると他のインスタンスタイプ・サイズと比べて安価で IOPS の値が高い.

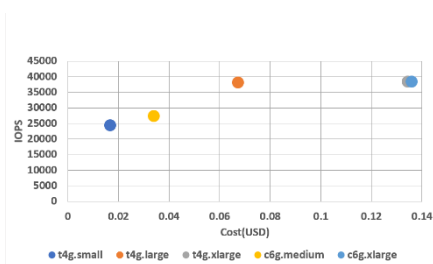


図 4 : Sequential-Read(SR)とコストの比較

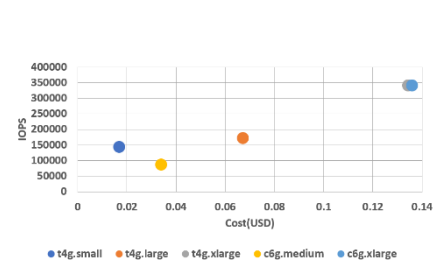


図 5 : Sequential-Write(SW)とコストの比較

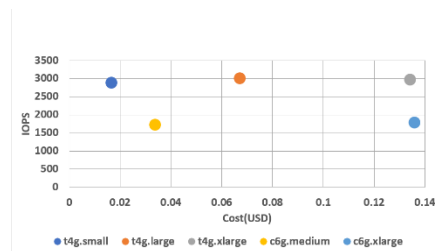


図 6 : Random-Read(RR)とコストの比較

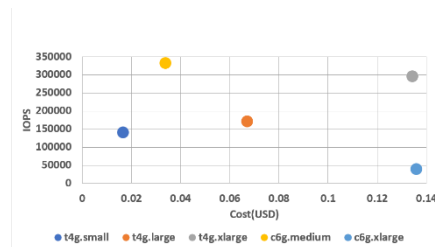


図 7 : Random-Write(RW)とコストの比較

4. 考察

本実験では 3 つの時間を各インスタンスタイプで計測し, インスタンスタイプの IOPS とコストを比較した. 時間の観点で考慮すると VDT においてコンピュータ最適化型である c6g が短い時間であった. VTT においては対照的に汎用型である t4g の方が短い傾向にある. TT においては t4g.large, c6g.xlarge の平均値が他のインスタンスタイプより, やや小さい. コストパフォーマンスの観点ではインスタンスタイプによって Sequential, Random または Read, Write の IOPS の値の差異が生じた. 今回の実験では 1 つのインスタンスに対してそれぞれ計測を行ったが科学技術計算ワークフローにおいては複数のインスタンスを同時に長時間利用するため, 汎用型かつコストが低いインスタンスタイプに対しても, 同様に評価し, 比較検討が必要となる.

参考文献

- [1] Ludäscher, Bertram, et al., Scientific workflow management and the Kepler system, Concurrency and Computation: Practice and Experience, Volume 18, Issue 10, pp.1039-1065, 2006.
- [2] S. Ostermann, A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema, A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing, Cloud Comp : Cloud Computing, pp.115-131, 2010.
- [3] Juve, G. Deelman, E. Berriman, G.B. et al., An Evaluation of the Cost and Performance of Scientific Workflows on Amazon EC2, Journal of Grid Computing, Volume 10, Issue 1, pp.5-21, 2012.