

D-03

クラウドコンピューティングにおけるコストを考慮した資源選択手法の検討 Resource Provisioning Algorithm using Cost Information in Cloud Environments

伊藤 耕平† 小坂 隆浩‡
Kohei Ito Takahiro Koita

1. はじめに

近年、クラウドコンピューティングが普及し、ユーザは手軽に大規模な計算資源を利用することが可能になった。Amazon が提供するクラウドコンピューティングサービスである Amazon EC2(Amazon Elastic Computing Cloud)[1]はユーザが利用料金を支払うことで、インターネットを通じて、CPU やメモリ、HDD などをもとまりにしたインスタンスと呼ばれる仮想計算機を利用することができる。計算資源は複数の種類が用意されており、それぞれ CPU やメモリなどの性能と単位時間あたりの料金が異なる。そのため、ユーザが同様のアプリケーションの使用を想定した場合でも、選択した計算資源によって処理時間や利用料金が異なる場合がある。例として、次のような場合がある。Amazon EC2 では CPU 性能を ECU という単位で表すが、CPU 性能が 1ECU であり利用料金が 1 ドル/時間であるインスタンス A と CPU 性能が 5ECU であり利用料金が 3 ドル/時間であるインスタンス B があるとする。図 1 のように、この 2 つの計算資源に 1ECU の CPU 性能で 10 時間かかる処理を実行した場合、インスタンス A は処理時間 10 時間で利用料金は 10 ドル、インスタンス B では処理時間は 2 時間で利用料金は 6 ドルとなり、同じ処理でも処理時間と利用料金に違いが生じる。

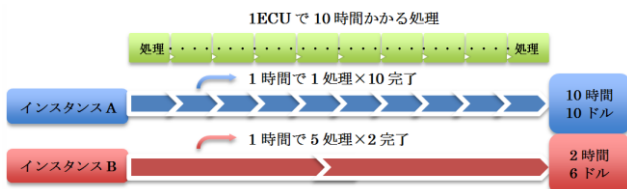


図 1 2つのインスタンスの比較

また、クラウドコンピューティングでは必要なときに必要なだけ計算資源を利用することができるという特徴から、大規模演算処理を行うことが注目されている。大規模演算処理では処理時間も長い。計算資源を持たないワシントン・ポストは AmazonEC2 を利用してサーバ 200 台相当の処理能力を確保することで、17000 ページを超える書類データの文字抽出処理を行った。この作業は 9 時間以内に完了し、利用料金は 144.62 ドルであったが、もし異なる計算資源を利用したとして、最終的に計算資源 1 台あたり 0.5 ドル利用料金が高くなったとする。1 台あたりの利用料金で比較すると少ない違いだが、200 台全ての利用料金の差は 100 ドルにも及んでしまう。そのため、クラウドコンピューティングにおける大規模演算処理においてはコストの影響が非常に大きい。

クラウドコンピューティングでは性能や利用料金の異なる様々な計算資源を利用することができる。ユーザの行いたい処理は一つであっても、どの計算資源を選択するかによって利用料金は異なる。ユーザが処理にかけることのできる利用料金は無限ではないため、できるだけ利用料金は低いほうが望ましい。しかし、計算資源の性能値や単位時間あたりの利用料金だけでは、処理に対する利用料金を導くことはできない。特に大規模演算処理においての計算資源の違いによる利用料金の差は大きくなる。そのため、クラウドコンピューティングにおいて適切な資源選択手法を検討し、利用料金を軽減させることが重要となる。

本研究では、クラウドコンピューティングサービスである AmazonEC2 上で大規模演算処理を行う場合を想定する。その際の利用料金に注目して、処理に対する利用料金の軽減を目的とし、資源選択手法を検討する。

2. Amazon EC2 の現状と課題

2.1 現状

AmazonEC2 は、インターネットを通じて仮想計算機を提供するクラウドコンピューティングサービスである。Amazon EC2 では提供される計算資源をインスタンスと呼び、インスタンスには複数の種類が用意されている。インスタンスのような計算資源が実際に存在しているデータセンタをリージョンと呼び、ユーザはリージョンを自由に選択することができる。インスタンスはハードウェア資源とソフトウェア資源をひとまとまりにしたものであり、ユーザはインスタンス上で自由に環境を構築できる。インスタンスはそれぞれ、性能値と一時間あたりの利用料金が設定されており、ユーザはどのインスタンスを利用するかを選択する。AmazonEC2 の各インスタンスの詳細を表 1 に示す。ここでのインスタンスの例は、リージョンが北カリフォルニア、OS が WindowsServer2008 の場合である。各インスタンスの CPU 性能は ECU という値で表され、1ECU は 1.0-1.2GHz 2007 Opteron, または 2007 Xenon プロセッサの CPU と同等の処理能力である。

表 1 各インスタンスの詳細

インスタンス	ECU	コア数	1 コアの ECU	メモリ (GB)	I/O 性能	利用料金 (\$/1h)
m1.small	1	1	1	1.7	標準	0.13
m1.large	4	2	2	7.5	高速	0.52
m1.xlarge	8	4	2	15	高速	1.04
m2.xlarge	6.5	2	3.25	17.1	標準	0.69
m2.2xlarge	13	4	3.25	34.2	高速	1.38
m2.4xlarge	26	8	3.25	68.4	高速	2.76
c1.medium	5	2	2.5	1.7	標準	0.31
c1.xlarge	20	8	2.5	7	高速	1.24

†同志社大学大学院 情報工学専攻, Graduate School of Informatics System Design, Doshisha University

‡同志社大学 理工学部, Faculty of Science and Engineering, Doshisha University

2.2 Amazon EC2 利用における問題点

AmazonEC2 の利用における問題点として、計算資源の性能と一時間あたりの利用料金だけでは、ユーザが行いたい処理に対する適切な計算資源の選択が困難であることが挙げられる。

Amazon EC2 では複数の計算資源が提供されており、ユーザは自由に計算資源を選択することができる。しかし、計算資源の選択にあたって、ユーザが判断する材料となるのは Amazon が示す計算資源の性能値と、1 時間あたりの利用料金だけである。例えば、図 2 のように、m1.small と c1.medium の 2 つの計算資源を比較した場合、1 時間あたりの利用料金は約 2.4 倍であるが、CPU 性能の差は 5 倍である。したがって、CPU 性能だけに依存する処理を行った場合、m1.small の処理時間は c1.medium の 5 倍かかる。よって、処理時間、利用料金ともに c1.medium の方が低くなる。

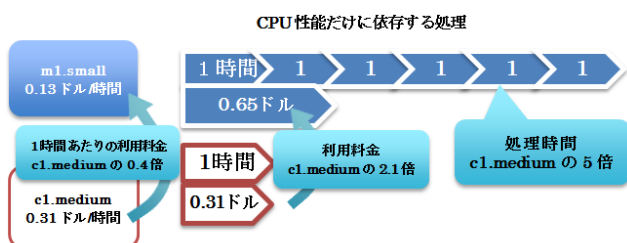


図 2 CPU 性能だけに依存する処理の計算資源の比較

このように、CPU 性能だけに依存するアプリケーションや処理であれば、上記のように性能値と 1 時間あたりの利用料金だけで、適切な計算資源を選択することができる。しかし実際のアプリケーションは CPU 性能だけでなく、メモリ性能や I/O 性能など、複数の処理性能の影響を受けるため、性能値と単位時間あたりの利用料金だけでは処理に対してどの計算資源が適切であるか判断することは難しい。処理を行う前に、処理と計算資源の関係性を把握することは困難であり、計算資源の選択によっては必要以上に利用料金を消費してしまう場合もある。特に、クラウドコンピューティングにおける大規模演算処理では、処理の実行前に計算資源のどの性能を必要とするのかを判断することが難しく、処理の経過とともに必要とする性能が変化することもある。そのため、計算資源の性能値と単位時間あたりの利用料金だけでは、行う処理に対してどの計算資源を選択するのか特定することが難しい。また、大規模演算処理の特徴として、処理時間が長いことが挙げられる。そのため、計算資源の違いによる利用料金の差は大きくなる。このような問題を解決するために、処理と計算資源の関係性を導き出すことで利用料金を考慮した資源選択手法を実現し、利用料金を軽減する必要がある。

3. 資源選択手法の検討

クラウドコンピューティングにおける資源選択手法では、計算資源の情報を取得し、負荷によって計算資源の数を増減させる資源選択手法が研究されている[2][3]。しかし、クラウドコンピューティングにおける資源選択手法に関する研究の多くは負荷分散が中心であり、利用料金については考慮されていない。

本研究では、利用料金を軽減するために、大規模演算処理においてどの計算資源を選択するかといった資源選択手法を検討する。大規模演算処理では、処理の内容と経過に

よって求める性能が CPU 性能からメモリ性能に変換するというような場合があり、それに応じて必要とする計算資源も変化する。そのため、処理に応じた計算資源を選択する必要がある。しかし、計算資源の性能値と単位時間あたりの利用料金だけでは、適切な計算資源を選択できない。資源選択手法を検討するためには、処理の内容と各計算資源の処理時間、利用料金や負荷情報の関係性について検討する必要がある。そのため、資源選択手法の検討における方針として、計算資源を評価するために予備実験を行い、各計算資源において利用料金、処理時間や負荷情報を評価する。計算資源の評価として、素数探索処理、行列演算処理、ファイル読み込み処理を各計算資源上で動作させ、利用料金と処理時間の評価を行う。この 3 つの処理はそれぞれ、Amazon が提示している性能値である CPU 性能、メモリ性能、I/O 性能に注目して計算資源を評価するための処理である。

計算資源の評価実験の環境として、Amazon EC2 の計算資源を用いる。Amazon EC2 のリージョンは、アジア・パシフィック(東京)リージョンを選択し、OS は Windows Server 2008 を使用する。評価のイメージを図 3 に示す。

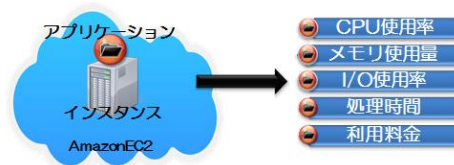


図 3 評価イメージ

予備実験から、CPU 性能、メモリ性能、I/O 性能の 3 つの負荷情報を必要とする処理における各計算資源の利用料金と処理時間を評価することができる。利用料金に着目して考えると、この 3 つの性能を求める処理に対して、それぞれ利用料金を軽減する見込みのある計算資源を導くことができる。

すなわち、計算資源の性能値と 1 時間あたりの利用料金だけでは得られなかった、どのような処理にどのような計算資源を選択すれば利用料金を軽減することができるという関係性を導くことが可能である。

4. まとめと今後の展望

本研究では Amazon EC2 上での大規模演算処理における利用料金の軽減を目的とし、資源選択手法を検討した。大規模演算処理では処理の経過に応じて求める性能が変化する場合があるため、予備評価で得られる単一の性能に依存する処理に対する各計算資源の評価に基づいた、処理の経過に対応した資源選択手法をさらに検討する必要がある。

参考文献

- [1] Amazon Elastic Compute Cloud (Amazon EC2) : <http://aws.amazon.com/jp/ec2/>.
- [2] 宮城亮太, 池部実, 猪俣敦夫, 藤川和利, 砂原秀樹, クラウド環境におけるサーバ負荷に応じた動的資源割当システムの提案と評価, 電子情報通信学会技術研究報告, 技術と社会・倫理 110(429), pp.17-22, 2011.
- [3] 豊島詩織, 山口実靖, 小口正人, 仮想マシン PC クラスターのロードバランスに関する評価と実クラウドへの適用の検討, DEIM フォーラム 2010 論文集, C2-1, 2010.