# パブリッククラウド環境での AutoScaling 機能の評価

1200334　土屋 椋　【 分散処理 OS 研究室 】

## 1　はじめに

サーバ負荷に応じて自動的にクラウドサーバのリソースを増減させる仕組みとしてオートスケールがある．オートスケールを用いると必要なサーバリソース量を自動で選択し，最適なサーバリソースでサーバを運用することができる．本研究ではパブリッククラウドの Amazon Web Services（以降，AWS と略す）が提供する AutoScaling 機能について評価する．オートスケールでサーバの処理性能を高める方法はスケールアップとスケールアウトの 2 種類存在する．AWS AutoScaling では，スケールアウトの仕組みを用いてサーバ全体の処理性能を管理する．

## 2　AWS AutoScaling 機能

AutoScaling は，Amazon CloudWatch が収集した AutoScaling グループの EC2 インスタンスのログを元に AutoScaling グループのインスタンス数を変化させる機能である．AWS AutoScaling は AWS マネジメントコンソール（以降，AWS-MC と略す）と AWS-SDK で使用できる．

AWS-MC では Amazon CloudWatch 上に，CPU 使用率が閾値を超えると発動するアラームを設定した．CloudWatch アラームが発動すると，スケールが変化する．AWS-SDK では Amazon CloudWatch が収集した AutoScaling グループの EC2 インスタンスの CPU 使用率を参照し，閾値を超えているとスケールを変化させるようプログラムした．

## 3　評価

インスタンスを 1 つ起動し，ロードバランサの URL に対して Apache Bench を用いて負荷をかける．負荷モデルとして，次の 2 種類のデータを用いる．

- 高知工科大学 Web サイトにおける入試結果発表時のアクセスデータ（図 1）
- EC サイトにおける SNS プッシュ通知時のアクセスデータ（図 2）[1]

これらの負荷モデルを用いて AutoScaling を行い，インスタンス数の増減遷移を確認した．インスタンスの起動には，パッケージのアップデート，Web アプリのビルド処理などで Web サーバとして稼働するまで約 5 分を要した．そのため，スケールアウトの閾値を低めに設定しスケールアウトの感度を高くした．

**スケールアウト**　直近 1 分の CPU 使用率が 30%以上時

**スケールイン**　直近 1 分の CPU 使用率が 10%未満時

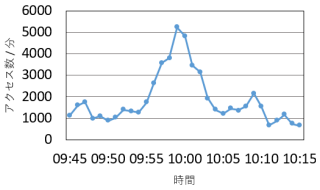


図 1　入試結果発表時

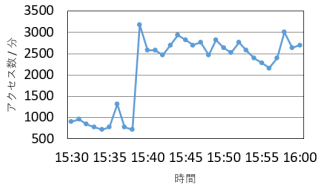


図 2　EC サイトのセール時

図 3，4 は AutoScaling によって変動したインスタンス数の遷移のグラフである．アクセスのピークである 10:00 頃のインスタンスの数は AWS-MC で設定した場合が 4，AWS-SDK で設定した場合は 5 となった．その後も AWS-SDK で設定したほうはインスタンスの減少が早い段階で発生しており，負荷に対して柔軟に対応できているといえる．

しかし，図 2 のような爆発的アクセスは AutoScaling だけでは間に合わず，サーバがリクエストをさばききることができなかった．そのため，大量のアクセスが発生することが明らかな場合には，事前にスケールアウトやスケールアップをしておくなどの対策が必要になる．

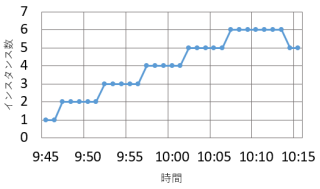


図 3　AWS-MC 設定時

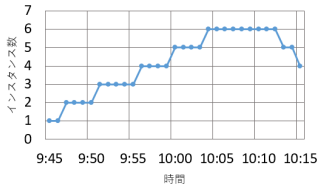


図 4　AWS-SDK 設定時

## 4　まとめ

本研究では、パブリッククラウドである AWS の AutoScaling 機能について実際の負荷モデルを元に評価した。本研究では、スケールの閾値に CPU 使用率を用いたが、AWS の AutoScaling 機能は、ネットワークの入出力状況を閾値にできる。それらを組み合わせることで、より柔軟なスケールが行えると考えられる。

## 参考文献

[1] 押田 知己，”負荷に弱い Web サイトはこうして落ちる! BtoC サイトに見るアクセス爆増（バースト）のパターンと備え・対策”，https://codezine.jp/article/detail/8793，2019 年 10 月閲覧．

[2] ”Amazon EC2 Auto Scaling とは”，https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/what-is-amazon-ec2-auto-scaling.html，2019 年 10 月閲覧．