

# コンテナキャッシュサーバを用いた分散 Web システムに関する研究

18T345 頼實 優樹 (最所研究室)

負荷量に応じてキャッシュサーバ数を動的に増減させる分散 Web システムが複数存在する環境において、動的に起動・停止を行い、アクセスを肩代わりするコンテナキャッシュサーバを導入したときの効果について述べる。

## 1 はじめに

近年のクラウド技術の発展により、クラウド上に仮想マシンとして構築したキャッシュサーバを分散 Web システムに用いることで、Web サービスの応答性を向上させることが容易になった。当研究室では、負荷量の増減に応じて、動的にキャッシュサーバを増加(スケールアウト)・減少(スケールイン)させることで応答性を確保しつつ、運用コストを低減することを目的とした分散 Web システムを開発している [1]。先行研究では、常時起動している共有キャッシュサーバを用いて、スケールアウト時のアクセスを肩代わりし、応答性の改善を図った。さらに、複数の分散 Web システムでそのサーバを共有し、常に起動することによってかかるコストの低減を図った。しかし、利用の有無にかかわらず、共有キャッシュサーバを維持するコストがかかり続けることは変わらない。本研究では、単一の分散 Web システムにおいて、共有キャッシュサーバを仮想マシンよりも起動処理のオーバーヘッドが少ないコンテナを用いてコンテナキャッシュサーバとして実装し、動的に起動・停止を行うことによる応答性の変化とさらなるコストの改善について述べる。

## 2 コンテナキャッシュサーバを用いる分散 Web システム

図 1 に、コンテナキャッシュサーバを用いた分散 Web システムの概要を示す。

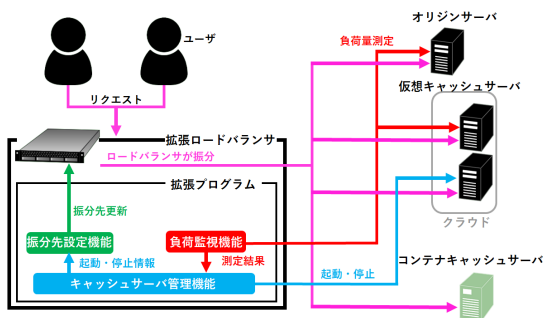


図 1: コンテナキャッシュサーバを用いた分散 Web システム

本システムは、先行研究で提案された共有キャッシュサーバを用いた分散 Web システムの共有キャッシュサーバをコンテナキャッシュサーバに置き換えたものであり、拡張ロードバランサ、オリジンサーバ、オリジン

ンサーバから取得したキャッシュを提供する仮想キャッシュサーバ群、動的に起動・停止を行い、スケールアウト時のアクセスを肩代わりするコンテナキャッシュサーバから構成される。

図 2 に、コンテナキャッシュサーバを用いたスケールアウト時の処理フローについて示す。

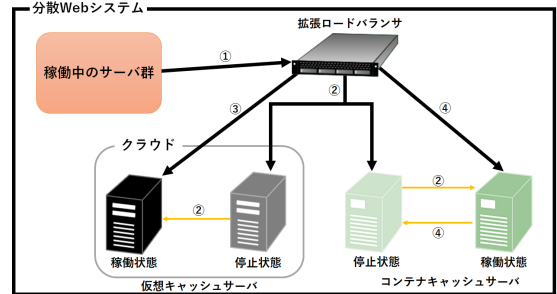


図 2: コンテナキャッシュサーバを用いたスケールアウト時の処理フロー

- ① 稼働中のサーバ群が過負荷になったことを拡張ロードバランサが検知する。
- ② 仮想キャッシュサーバの起動処理と並行して、コンテナキャッシュサーバの起動処理を行う。
- ③ コンテナキャッシュサーバの起動処理完了後、コンテナキャッシュサーバへのリクエスト振分を開始する。
- ④ 仮想キャッシュサーバの起動処理完了後、仮想キャッシュサーバへのリクエスト振分を開始する。
- ⑤ 規定時間後にコンテナキャッシュサーバへのリクエスト振分を停止し、処理中のリクエストの処理が完了すると、コンテナの停止処理を行う。

## 3 負荷実験と考察

単一の分散 Web システムにおいて、コンテナキャッシュサーバを用いた分散 Web システムと共有キャッシュサーバを用いた分散 Web システム、オリジナルの分散 Web システムで比較実験を行った。実験に用いた各サーバは、仮想マシンを用いて図 3、図 4 の性能で構築した。コンテナキャッシュサーバは仮想マシン上でコンテナを構築している。拡張ロードバランサ上で、ソフトウェアロードバランサと拡張プログラムを起動する。次に、各クライアントサーバからの秒間

リクエスト数を最大値である 500 まで規定時間かけて増加させる。そして、100 秒間最大値リクエスト数を維持し、その後秒間リクエスト数を規定時間かけて減少させていく。負荷量の増加率は必ずしも一定ではないため、規定時間を 60 秒、120 秒、240 秒の 3 つのパターンで設定し実験を行った。実験結果から得られた平均応答時間と負荷量を規定時間別で図 5、図 6、図 7 に示す。各規定時間での実験結果において、負荷量はオリジナルの分散 Web システム、コンテナキャッシュサーバを用いた分散 Web システム、共有キャッシュサーバを用いた分散 Web システムの順に高くなっていく。そして、どの規定時間で実験を行った場合でもコンテナキャッシュサーバを用いることでオリジナルの分散 Web システムより応答性が上がっている。また図 5 から、設定した規定時間が短いほど共有キャッシュサーバを用いた場合の平均応答時間と負荷量に大きな差があることが読み取れる。一方で図 6 と図 7 からは、規定時間が長いほど平均応答時間と負荷量に大きな差は生まれていないことが読み取れる。このことから、コンテナキャッシュサーバを用いた分散 Web システムは、起動時間を抑えることでコストを削減しつつ、オリジナルの分散 Web システムよりも応答性が高い。さらに、負荷量の上昇率によっては共有キャッシュサーバを用いた分散 Web システムに近い応答性を発揮することができる。

#### 4 まとめ

コンテナキャッシュサーバを用いた分散 Web システムを開発し、単一 Web システム上で先行研究で提案された 2 種の分散 Web システムとの比較実験を行った。オリジナルの分散 Web システムより高い応答性があり、負荷量の上昇率によっては共有キャッシュサーバを用いた分散 Web システムに近い応答性を発揮することを確認した。今後の課題として、負荷量が急激に増加した場合の対策や複数の Web システム上での評価実験、Azure 上での負荷実験がある。

#### 参考文献

- [1] 河野彰吾, 最所圭三, “キャッシュサーバを用いる分散 Web システムにおける応答性の改善”, 第 18 回情報科学技術フォーラム講演論文集, L-002, Vol.4, pp.111-114, 2019

役割	CPU コア数 [コア]	メモリ量 [GB]
拡張ロードバランサ 1 台	2	4
オリジンサーバ 1 台	1	2
仮想キャッシュサーバ 4 台	1	2
クライアント 2 台	1	2

図 3: 共通の仮想マシンの性能

役割	CPU コア数 [コア]	メモリ量 [GB]
共有キャッシュサーバ 1 台	1	2
コンテナキャッシュサーバ 1 台	1	2

図 4: 異なる仮想マシンの性能

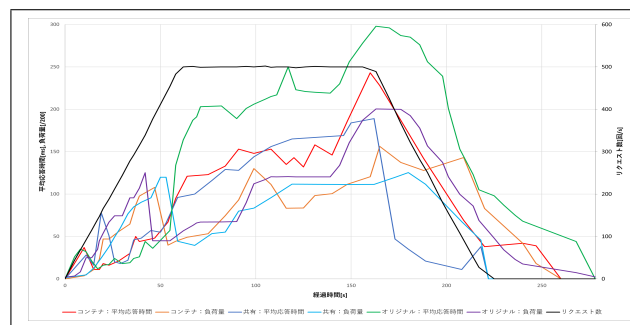


図 5: 規定時間 60 秒の実験結果

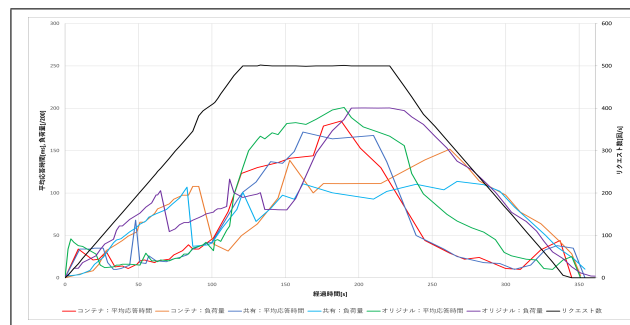


図 6: 規定時間 120 秒の実験結果

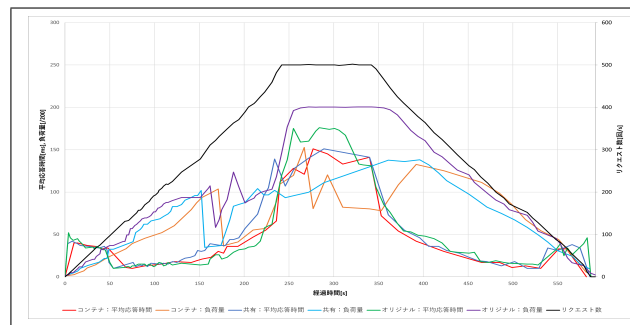


図 7: 規定時間 240 秒の実験結果