

УДК 004.04

Исследование среднего потребления CPU и RAM высоконагруженных и слабнонагруженных сервисов в течение времени с использованием Knative
Рощин М.В. (Университет ИТМО), **Шалгуева С.Л.** (Санкт-Петербургский политехнический университет Петра Великого)
Научный руководитель – доцент, кандидат технических наук, Штенниев Д.Г.
(Университет ИТМО)

Введение. В последние несколько лет в облачных технологиях появилась возможность использования бессерверных технологий. [2] Безусловно отличный инструмент позволяющий сэкономить на потребляемых ресурсах. Множество сервисов предоставляют возможность использования бессерверных вычислений, такие как Yandex, Amazon, Azure, Google. Многие компании уже внедряют serverless технологии в свои продукты. Однако до конца не ясно, где находится точка, после которой использование бессерверных технологий уже не приносит выгоду, а наоборот начинает потреблять больше денежных ресурсов чем обычный сервис. В нашем исследовании производится сравнение среднего потребления CPU и RAM сервисами с разной нагрузкой в течение времени на базе self-hosted фреймворка Knative [1].

Основная часть. Для исследования были выбраны 4 системы с разной степенью нагруженности.

- 1) Сервис «А» с нагруженностью в 90%, потреблением CPU 1 и RAM 2GB
- 2) Сервис «В» с нагруженностью в 10%, потреблением CPU 1 и RAM 2GB
- 3) Сервис «С» с нагруженностью в 1%, потреблением CPU 1 и RAM 2GB
- 4) Сервис «D» с нагруженностью в 0.1%, потреблением CPU 1 и RAM 2GB

Для каждого из перечисленных выше сервисов были построены графики отражающее потребление среднее ресурсов в зависимости от количества сервисов каждого типа. В рамках исследования кол-во сервисов каждого типа увеличилось до тех пока не выстраивалось четкая картина движения графика с использованием Knative и без его использования.

Таким образом в дальнейшем каждый график был аппроксимирован и была найдена точка эффективного использования технологии Knative для каждого из типов сервисов. Графики построены отдельно как для среднего потребление RAM, так и для среднего потребления CPU в зависимости от количества сервисов каждого типа.

Особенности расчета эффективности использования бессерверных вычислений в данном случае заключается в том, что Knative сам по себе потребляет определенное кол-во ресурсов, отличающейся в зависимости от того с каким кол-вом нод он работает. Тем самым делается очевидным факт того, что чем больше на облаке находится сервисов, тем более эффективным будет использование Knative как технологии контроля автоскалирования сервисов из 0 и в 0.

Однако главным вопросом всё также остается необходимость определить точку, после которой использование данной технологии дает преимущество по сравнению с обычным облачным сервисом не контролируемым автоскалированием.

Выводы. Проведен анализ среднего потребления CPU и RAM в различных системах и выявлена зависимость эффективности использования в зависимости от нагруженности сервиса. Для сервисов типа «А» использование Knative оказалось не рациональным, так как использование данной технологии лишь добавляет дополнительную нагрузку и не несет никаких улучшений. Для сервисов типа «В» точка эффективного использования наступает при 10 сервисах для RAM и при 20 сервисах для CPU. Для сервисов «С» и «D» точка эффективности наступает при 5 сервисах для RAM и также при 20 сервисах по CPU. Такое число эффективности для CPU объясняется тем, что Knative сам по себе потребляет ресурсы и нельзя достичь большей эффективности использования, т. к. Knative потребляет все выделенные под развертывания ресурсы.

Список использованных источников:

1. Towards Serverless as Commodity: a case of Knative // ResearchGate URL: https://www.researchgate.net/publication/336567672_Towards_Serverless_as_Commodity_a_case_of_Knative (дата обращения: 01.02.2023).

2. Особенности технологий бессерверных вычислений // КиберЛенинка URL: <https://cyberleninka.ru/article/n/osobennosti-tehnologiy-besservernyh-vychisleniy> (дата обращения: 15.01.2023).