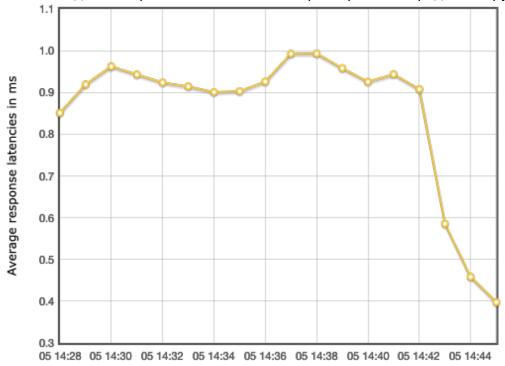
Отчёт о влиянии репликации на производительность.

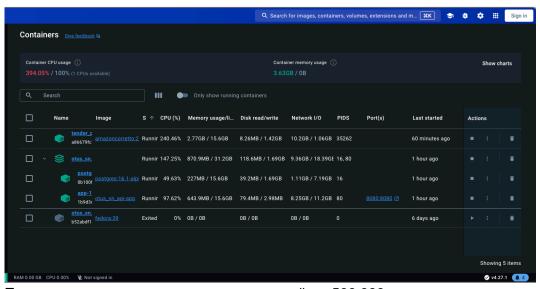
В ходе эксперимента по репликации бд были протестированы два эндпоинта: /user/register и /user/{id}. Кластер состоял из одного примари и двух реплик работающих в синхронном режиме.

Едпоинт /user/{id}

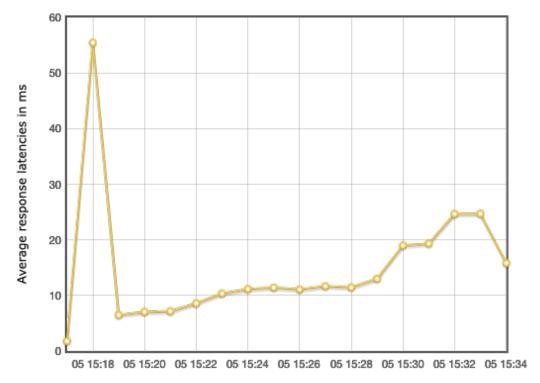
При 250 000 пользователях по 10 запросов каждый, среднее время отклика составляет 0,9 мс и 1 мс для 99 персентиля, ошибки не зафиксированы. Средняя загрузка CPU - 40-50%.



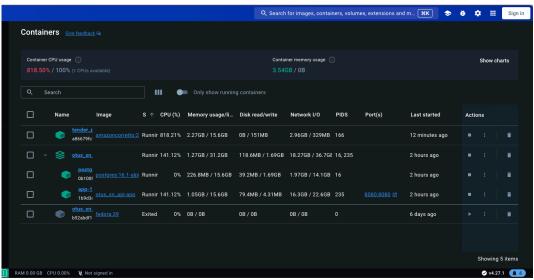
Elapsed Time (granularity: 1 min)



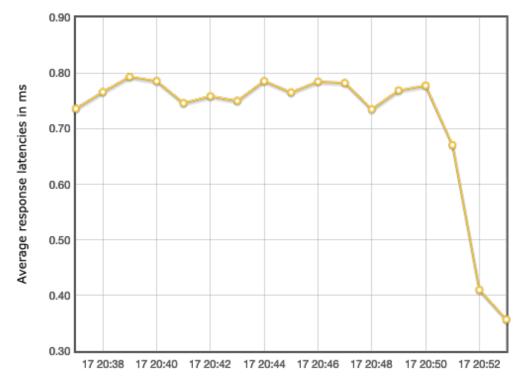
При увеличении количества пользователей до 500 000 производительность значительно падает, загрузка CPU составляет 100%, количество ошибок составляет 11,6%. Среднее время отклика - 21,6 мс и 170 мс для персентиля 99.



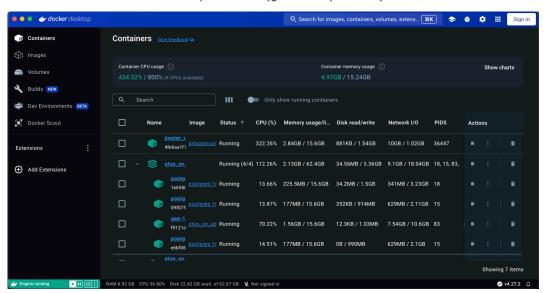
Elapsed Time (granularity: 1 min)



После перевода бд в режим синхронной репликации с одним примари и двумя репликами, общая производительность заметно не изменилась, CPU остаётся ограничивающим фактором. Но распределение ресурсов между узлами стало более пропорциональным.

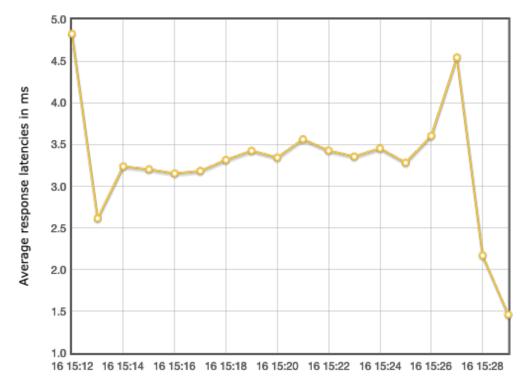


Elapsed Time (granularity: 1 min)

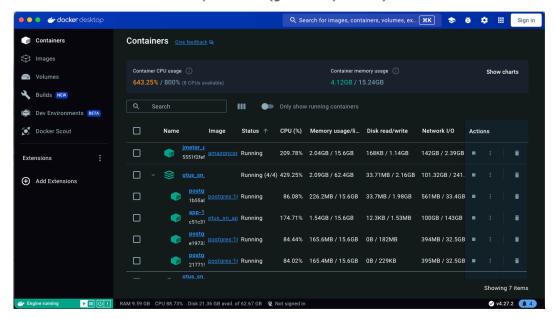


Ендпоинт /user/search

Производительность данного ендпоинта без репликации описана в общем отчёте по производительности. При активации репликации данный ендпоинт ведёт себя аналогично предыдущему. Среднее время отклика составляет 3,5 мс и достигает 4мс для 99 персентиля. Загрузка CPU составляет около 80 процентов.



Elapsed Time (granularity: 1 min)



Выводы.

Во всех тестах ограничивающим фактором является СРU. Так как все экземпляры запускались на одной физической машине под докером, увеличение количества экземпляров при активации репликации увеличило количество ресурсов потребляемых виртуально машиной докера и не позволило полностью раскрыть весь потенциал репликации для запросов на чтение.