

Задание №7 в рамках вычислительного практикума.

Постановка замерного эксперимента

Кострицкий А. С., Ломовской И. В.

Москва — 2022 — TS2204242039

На основе задачи №5 ЛРН№2 по курсу «Программирование на Си» проведите сравнение производительности работы программы по двум плоскостям:

1. Разные способы работы с элементами одномерного массива:
 - (a) использование операции индексации `a[i]`;
 - (b) формальная замена операции индексации на выражение `*(a + i)`;
 - (c) использование указателей для работы с массивом.
2. Разные уровни оптимизации: `Os`, `O0`, `O1`, `O2`, `O3`.

Иначе говоря, при компиляции должны получиться $3 \times 5 = 15$ вариантов программы. Для сравнения производительности следует реализовать несколько скриптов:

1. `build_apps.sh`, вызвав который, можно получить весь набор необходимых исполняемых файлов.
2. `update_data.sh`, вызвав который, можно добавить некоторые данные в датасет экспериментов.
3. `make_preproc.sh|py`, вызвав который, можно подготовить данные из набора, провести первичный анализ: посчитать среднее арифметическое, медианное, найти максимум и минимум, вычислить нижний и верхний квартили.
4. `make_postproc.sh|py`, вызвав который, можно получить указанные ниже графики.
5. `go.sh`, вызвав который, можно получить данные эксперимента (скрипт вызывает по очереди предыдущие четыре).

В отчёте привести следующие **графики**:

1. Обычный кусочно-линейный график зависимости времени выполнения в любых единицах измерения времени от числа элементов массива для всех 15 вариантов программы.
2. Кусочно-линейный график с ошибкой (среднее, максимум, минимум) для всех вариантов обработки массива при уровне оптимизации `O2`.
3. График с усами (среднее, максимум, минимум; нижний, средний и верхний квартили) для варианта обработки «через квадратные скобки» при уровне оптимизации `O3`.

В отчёте привести **таблицы**:

1. Таблицы ко всем указанным графикам. Напоминаем, что график, если не имеется цель лишь схематично изобразить что-либо, всегда рисуется по табличным данным.
2. Таблицу для результатов обработки «через квадратные скобки» с уровнем оптимизации 02 со столбцами: длина массива n , время выполнения t_n , величина $\frac{\ln(t_{i+1}) - \ln(t_i)}{\ln(n_{i+1}) - \ln(n_i)}$ для всех строк, кроме последней.

Примечания:

1. Помните, что для уменьшения влияния побочных эффектов время замеряют только у целевого алгоритма.
2. Проводить эксперимент следует со массивами размером от минимального (в некоторых задачах это 1, в некоторых – несколько больше) до, как минимум, 10000. Если производительность Вашего ПК позволяет собрать результаты 20 тестов в час для массивов большего размера, можно поместить эти результаты в отчёт. Если производительность ПК не позволяет на 10000 элементов собирать хотя бы 20 результатов в час, то максимальный размер обговаривается отдельно.
3. Шаг измерений – 500. Можно плотнее.
4. Помните, что для чистоты эксперимента состояние системы не должно зависеть от того, какой конкретно алгоритм был использован. Финальное измерение времени всегда проводится тогда, когда система приведена к ожидаемому состоянию.
5. Для отрисовки графиков можно использовать gnuplot, matplotlib или любой другой «лёгкий» пакет. Два требования: возможность нарисовать всё, что требуется в задании, и наличие консольного интерфейса.
6. При желании разрешается добавить другие плоскости исследования: наличие/отсутствие `restrict`, различные максимальный и текущий размеры статического массива, etc.

В отчёте объясните полученные результаты и приведите ответы на следующие **вопросы**:

1. Какой способ обработки быстрее и почему?
2. В датасете появился статистический выброс, причины которого очевидны, например, эксперимент был поставлен на другой машине. Порядок экспериментов не был известен заранее. Можно ли вырезать данные со статистическим выбросом из датасета?
3. В датасете появился статистический выброс. Известно, что эксперимент был поставлен вчера, существует резервная копия датасета за позавчера. Можно ли вырезать данные со статистическим выбросом из датасета?
4. В датасете обнаружена серия экспериментов с одним результатом. Можно ли заменить её одним экспериментом?
5. Если заполнение случайными числами массива (или любая другая инициализация) присутствует в каждом эксперименте, то почему Вы замеряете время только у целевого алгоритма?