

ПРИМЕНЕНИЕ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ В РЕШЕНИИ ТРУДОЁМКИХ ЗАДАЧ ПО ОБРАБОТКЕ ДАННЫХ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Рымашевский О.С.

Куликов С.С. – к.т.н., доцент

В работе разработан алгоритм распределения вычисления произвольных пользовательских программ между несколькими вычислительными узлами, представляющими собой обычные персональные компьютеры. Для обеспечения возможности разделения задачи на независимые подзадачи пользовательская программа должна иметь строго заданную структуру. В докладе приводится обоснование решений, принятых в ходе разработки системы обеспечения распределённых вычислений.

В настоящее время в различных сферах часто возникает необходимость в расчёте вычислительно сложных задач. К таким задачам можно отнести следующие: цифровая обработка и преобразование аналоговых сигналов, распознавание графических образов в видеопотоке, обработка больших объёмов данных. Известными проектами, решающими задачи такого рода, можно назвать Folding@home (проект, моделирующий свёртывание молекул белка), SETI@home (проект, анализирующий данные с радиотелескопа обсерватории Оресибо на предмет искусственных сигналов), Climateprediction.net (проект, прогнозирующий изменения климата Земли в ближайшие 50 лет). Эти проекты рассчитаны на использование простаивающих вычислительных мощностей ПК обычных пользователей для полезных вычислений.

Статистика показывает, что добровольные распределённые компьютерные сети имеют значительно больший потенциал и большие экономические выгоды, чем отдельные высокопроизводительные вычислительные машины. Например, проект Folding@home имеет в распоряжении «сырую» вычислительную мощность всего в полтора раза меньше, чем мощнейший суперкомпьютер на сегодняшний день (20.7 petaFLOPS против 33.86 petaFLOPS у Тяньхэ-2).

Таким образом, задача создания программного средства, организующего распределённые вычисления, является актуальной и востребованной для улучшения исследовательской деятельности; также организацию распределённых вычислений можно считать полезной обучающей задачей.

Предлагаемая система распределённых вычислений принципиально отличается от перечисленных выше. Основное отличие – другая направленность вычисляемых задач. Перечисленные проекты решают одну сложную задачу на огромном объёме данных. Предлагаемая система проектируется универсальной, направленной на вычисление произвольного количества различных пользовательских задач.

Вычисляемые задачи задаются пользователем на скриптовом языке программирования Python, и использование скриптового языка программирования можно назвать вторым по важности отличием. Выбранный язык программирования предполагает упрощение разработки отдельных задач за счёт скорости их вычисления. Однако, использование JIT-компилятора, встроенного в интерпретатор PyPy или Stackless, позволяет значительно улучшить эффективность математических вычислений на Python.

Также могут задаваться дополнительные параметры вычисляемых задач: мера избыточности вычислений (количество узлов, вычисляющих одну подзадачу с одинаковыми входными параметрами), критерий выбора результатов вычисления подзадачи (в случае различных результатов одной подзадачи, присланных разными вычислительными узлами), поведение в случае непредвиденных ошибок в процессе вычисления подзадачи.

Преимущества предлагаемой системы распределённых вычислений состоят в простоте использования как клиентской, так и серверной части приложения. Система может работать на любой платформе, для которой есть реализация интерпретатора Python. Для запуска клиента предлагаемой системы достаточно запустить приложение с помощью PyPy и указать адрес сервера. Для запуска серверной части достаточно указать адрес сервера с базой данных приложения.

Предлагаемые программные решения и алгоритмы могут быть использованы для моделирования таких задач как:

- Исследование и оптимизация алгоритмов обработки больших объёмов данных.
- Реализация алгоритмов параллельных и распределённых вычислений.
- Реализация алгоритмов функционирования сложных гетерогенных распределённых вычислительных систем.
- Исследование показателей качества систем распределённых вычислений.

Предлагаемые решения могут быть использованы равно как для самостоятельных исследований, так и для демонстрации клиент-серверного подхода к организации распределённых сетей.