

Обработка нестационарного потока запросов в гетерогенных распределенных вычислительных системах

И. В. Логинов, Е. В. Лебедеенко

В крупных научных и промышленных организациях для решения задач текущей деятельности широко применяются различные вычислительные задачи, требующие значительных ресурсов. Обработка таких задач происходит на основе вычислительной инфраструктуры организации, основой которой является распределенная вычислительная система (РВС). Задачи для обработки формулируются в виде запросов к узлам РВС, обрабатывающие требуемый вид запросов. Запросы представляют собой специальным образом сгенерированные задачи различных областей науки и техники. В качестве таких задач могут выступать: комплексного моделирования сложных элементов корпоративных ИВС в распределенных гибридных инструментальных комплексах в процессе сопровождения АСУП; обработки результатов экспериментов [1]; имитационного моделирования объектов реального мира; обработки данных дистанционного зондирования земли, том числе в геоинформационных системах; обработка статистической информации и поддержка принятия решений, в том числе в OLAP-системах [1].

Имеется РВС, состоящая из разнородных узлов, соединенных различными линиями заданной пропускной способности. На вход системы поступает нестационарный поток запросов от различных пользователей. Параметры запросов, такие как время поступления, ресурсоемкость, и другие варьируются в широких пределах. Каждый запрос обладает весовой функцией $\xi(t)$ определяемой пользователем [2], сгенерировавшим запрос, на основе выделенного ему лимита. Весовая функция является монотонно убывающей от времени. Требуется разработать такой способ обработки нестационарного потока запросов в ГРВС, который будет максимизировать качество обслуживания пользователей, определяемое целевой функцией следующего вида:

$$Z = \sum_{i=0}^n \xi_i^Q(t_i - T_i), \quad (1)$$

где n - количество обработанных запросов за период, T - момент выполнения запросов, t - момент поступления запроса, Q - запрос. В целях максимизации целевой функции (1) предлагается использовать стратегию планирования, основанную на прогнозировании ресурсоемкости обработки запросов. Система планирования строится по распределенному принципу. Общая схема планирования обработки запросов включает в себя следующие операции: прием запроса со значениями параметров, заданными пользователем; классификация запроса и прогнозирование его ресурсоемкости на основе модели запросов класса; классификация вычислительных узлов ГРВС; объединение их в группы по классам и упорядочивание по удельной производительности; определение плана вычислений на основе спрогнозированной ресурсоемкости запроса и его весовой функции с целью достижения максимума целевой функции (1) на имеющихся в доступности вычислительных узлах ГРВС, путем использования алгоритма стохастической оптимизации.

Направлениями дальнейших исследований являются: разработка полнофункционального прототипа, реализующего предложенный способ в полном объеме; адаптация различных методов оптимизации под решение предложенной задачи и их сравнительный анализ на различных входных данных; определение эффективности применения предложенного способа при различных входных данных.

Литература

1. Воеводин, В. В. Параллельные вычисления [Текст] / В. В. Воеводин, Вл. В. Воеводин. – СПб.: БХВ-Петербург, 2002. – 608 с.: ил.
2. Бредихин, С.В. Распределение процессорного времени и памяти с помощью цен / С. В. Бредихин, И. А. Вялков, А. Б. Хуторецкий // Вычислительные технологии. 2006. Т. 11. С. 102-113.