**Краткие отчеты по проекту РФФИ №16-37-00240-а**

**Название проекта:**

**«Разработка и проведение исследований информационно-вычислительной технологии анализа генетических данных на гибридной супер-ЭВМ».**

**Руководитель – к.т.н. Якименко А.А.**

**2016 год**

Спроектирована параллельная архитектура программного обеспечения для высокопроизводительного генетического анализа на гибридной супер-ЭВМ. Использованы подходы проектирования параллельного программного обеспечения, методы адаптации программного кода, алгоритмов на архитектуры вычислительных систем на графических ускорителях.

Разработан высокопроизводительный параллельный программный комплекс для ускоренного выполнения перестановочного теста на гибридном суперкомпьютере НКС-30Т+GPU. Использованы подходы проектирования программного обеспечения для создания двухуровневого распараллеливания MPI+CUDA. Результаты численных экспериментов в задачах детерминации генетических признаков, полученные на эмпирическом материале, представленные как в виде данных по достигнутому ускорению, эффективности, масштабированию, так и в виде данных, обладающих научной значимостью и новизной. Использованы методы статистического анализа для определения статистически значимых величин, технологии моделирования статистических процессов.

Реализован инновационный модуль программного комплекса для проведения множественного эксперимента внутри одного перестановочного теста, позволяющий выполнять пакетную обработку данных однотипных экспериментов для сокращения накладных расходов времени запуска. Применен подход, позволяющий агрегировать результаты эксперимента в формат представления данных для проведения расчетов на гибридных высокопроизводительных архитектурах.

Исполнитель – Грищенко Михаил Владимирович (аспирант)

**2017 год**

Выполнено исследование производительности разработанного информационно-вычислительного комплекса путем проведения моделирования (Рис.1, 2). Проведено несколько моделирований с различными входными параметрами.

В результате при низкой загруженности системы, в основном она находится в режиме простоя, что для данной системы является стабильным результатом. Но при большой загруженности, которая не предполагается, но вероятна, коэффициент загруженности вычислительного кластера равен 1, из-за чего систему нельзя назвать эффективной – все задачи простаивают в очереди.



Рис. 1. Количество задач, отправленных на сервер и кластер с течением времени



Рис. 2. Количество задач в очереди на кластер, и количество задач, выполняющихся на сервере

После проведения полного факторного эксперимента были выявлены значимые факторы, которыми оказались: среднее время поступления задачи и процент задач с ошибкой в входных данных. Если время поступления задач в реальных условиях нельзя уменьшить, то процент ошибки в исходных данных можно избежать, валидируя данные перед отправлением на вычислительный кластер. Если система будет загружена, необходимо рассмотреть возможность включения дополнительных серверов (возможно виртуальных) для решения задач в автоматическом режиме или администратором используя информационную систему.

Исходя из всего вышеперечисленного можно сделать вывод о том, что разработанная система обладает широкими возможностями для дальнейшего развития с целью увеличения производительности.

Спектр требований к созданной технологии растёт в сторону повышения производительности вычислений. Решением данной проблемы является возможность включения дополнительных вычислительных мощностей для решения задач в автономном режиме или администратором, используя информационную систему.

Кроме этого существует необходимость в популяризации данной системы. Этого можно добиться путём добавления других статистических программ, которые специалисты смогут использовать для расчетов.