

## Участие облачных инфраструктур ОИЯИ и его стран-участниц в исследовании COVID-19 проекта Folding@Home

*И. Пелеванюк, Н. Кутовский*

В декабре 2019 года весь мир впервые услышал о новом инфекционном заболевании COVID-2019. Всемирная Организация Здравоохранения (ВОЗ) 20-го марта назвала сложившуюся ситуацию пандемией. Сегодня, к концу июля, общее количество учтённых заболевших составляет более 15 миллионов человек. Исследования, направленные на изучение структуры вируса SARS-CoV-2, оказались в фокусе мирового научного сообщества.

Объединённый Институт Ядерных Исследований (ОИЯИ) с апреля 2020 года принимает участие в изучении вируса SARS-CoV-2 в рамках проекта Folding@Home. Изначально, использовались ресурсы грид-сайтов, обрабатывающих данные, полученные с Большого Адронного Коллайдера (БАК). Задачи на грид-ресурсы рассылаются коллаборациями ATLAS, ALICE и CMS, в том числе и на грид-инфраструктуру в ОИЯИ.

С мая 2020 г. в работу включили свободные от задач облачные ресурсы ОИЯИ и его стран-участниц. На облаках за два месяца было выполнено более 5800 задач, из которых половина выполнена на облаке ОИЯИ, а другая половина – на облаках следующих организаций:

- Российский экономический университет им. Г.В. Плеханова,
- Институт ядерных проблем Белорусского государственного университета,
- Северо-Осетинский государственный университет им. К.Л. Хетагурова,
- Институт физики Азербайджанской национальной академии наук,
- Институт ядерных исследований и ядерной энергетики Болгарской академии наук.

Понимание структуры белков SARS-CoV-2 крайне важно, так как может ускорить открытие новых терапевтических методов лечения COVID-19. Изучение структуры белков ведётся в том числе с помощью методов молекулярной динамики. Они позволяют изучать весь спектр состояний, которые принимает белок, но для реализации этих методов требуются значительные вычислительные ресурсы. Современные наборы данных в основном содержат несколько микросекунд симуляции одного единственного белка. Некоторые наборы достигают длительности в миллисекунды. Но многие процессы в белках происходят на более длительных промежутках времени. Таким образом, симуляция всех белков, связанных с SARS-CoV-2 с длительностью, достаточной для их биологического анализа, требует беспрецедентного объёма вычислительных ресурсов.

Для решения этой задачи более миллиона людей и организаций по всему миру выступают волонтерами, предлагая вычислительные ресурсы своих компьютеров для проведения симуляций белков SARS-CoV-2. Работа такой коллаборации стала возможна благодаря платформе для распределённых вычислений Folding@Home, которая в 2020 году преодолела Эксафлопсный барьер и стала самым большим распределённым суперкомпьютером в мире. Используя этот ресурс, были получены наборы данных длительностью в десятые доли секунд. Эти данные, подтверждённые разнообразными экспериментальными наблюдениями, уже сегодня размещены в открытом доступе на сайте <https://covid.molssi.org/>.

Folding@Home - это общество волонтеров, исследователей и организаций, которые своими интеллектуальными и вычислительными ресурсами помогают понять динамику белков, их функции и дисфункции, с целью нахождения новых белков и лекарств. Проект был основан в 2000 году для изучения сворачивания белка. В те времена симуляция сворачивания даже маленьких белков на одном компьютере могла с лёгкостью затянуться на тысячи лет. Для того, чтобы справиться с этим вызовом, Folding@Home разделил одну огромную задачу на

множество маленьких симуляций - рабочих блоков, которые могут быть выполнены независимо друг от друга.

Спустя годы, проект эволюционировал и расширил спектр решаемых проблем, связанных с раком, сопротивляемостью микробам и болезням, связанным с дисфункцией белковых мембран. С появлением SARS-CoV-2 Folding@Home сменил фокус. Множество людей и организаций по всему миру нашли способ действовать, не смотря на обстоятельства, в которых можно было почувствовать себя беспомощным. Менее чем за три месяца количество активных вычислительных устройств в системе выросло с 30000 до одного миллиона. В работу включились не только волонтеры, но и организации: NVIDIA, Amazon, CERN, Intel и многие другие смогли выделить свободные ресурсы для задач Folding@Home.

ОИЯИ и его страны-участницы решили внести свой вклад. Изначально он делался в рамках участия в обработке данных Большого Адронного Коллайдера. Основная идея заключалась в использовании свободных вычислительных ресурсов грид. С идеей использования свободных ресурсов, задачи Folding@Home были запущены на облака ОИЯИ и его стран-участниц. Спустя два месяца работы на облаках ОИЯИ и его стран-участниц было выполнено 5800 задач.

Приоритетами во время запуска этих задач стало следующее: исключение влияния на виртуальные машины, уже работающие на облаках; возможность оперативно убрать нагрузку Folding@Home; сбор данных о количестве работы, выполненной на каждой из облачных инфраструктур; высший приоритет задачам изучения SARS-CoV-2.

Ресурсы распределённой облачной инфраструктуры ОИЯИ и его стран-участниц объединены в единую среду посредством промежуточного программного обеспечения DIRAC Interware. Использование этой платформы позволило значительно сократить работу, связанную с управлением задачами, и дало полный контроль над их запуском и остановкой.

Для учёта предоставленных ресурсов в Folding@Home была создана группа "Joint Institute for Nuclear Research", объединяющая отдельные учётные записи каждого из участников распределённой облачной инфраструктуры ОИЯИ. Потреблённый конкретной задачей ресурс засчитывается в качестве вклада соответствующей облачной инфраструктуры и группы JINR. Благодаря этому, на странице статистики Folding@Home можно узнать точное количество задач, выполненных на разных облаках, а также суммарное количество задач, выполненное всей группой.

Спустя два месяца работы на облаках ОИЯИ и его стран-участниц было выполнено порядка 6 тысяч задач, а общий объём работы в CPU часах даже в два раза превысил вклад ОИЯИ, сделанный под флагом эксперимента ALICE. Запущенные в облака задачи Folding@Home позволили получить информацию о реальной производительности разных облачных инфраструктур. Это дало возможность выявить узкие места инфраструктуры и их устранить.

Работа, выполненная всеми облаками, интегрированными в DIRAC:

<https://stats.foldingathome.org/team/265602>

Работа, выполненная CERN:

<https://stats.foldingathome.org/team/38188>

Вклад ОИЯИ совместно с экспериментом ALICE:

<https://stats.foldingathome.org/donor/84204021>

Вводная часть и описание во многом переведены из оригинальной статьи:

Citizen Scientists Create an Exascale Computer to Combat COVID-19

<https://www.biorxiv.org/content/10.1101/2020.06.27.175430v1.full.pdf> .