Participation of cloud infrastructures of JINR and its Member States in the COVID-19 study of the Folding@Home project

I. Pelevanyuk, N. Kutovskiy

In December 2019, the whole world first heard about a new infectious disease COVID-2019. On 20 March 2020, the World Health Organization (WHO) called the current situation a pandemic. Today, by the end of July, the total number of registered cases is more than 15 million. Investigations aimed at studying the structure of the SARS-CoV-2 virus have become the focus of the world scientific community.

The Joint Institute for Nuclear Research (JINR) has been taking part in the study of the SARS-CoV-2 virus within the Folding@Home project since April 2020. The resources of grid sites, which process data acquired from the Large Hadron Collider (LHC), were originally used. Work units are sent to grid resources, including the grid infrastructure of JINR, by the ATLAS, ALICE and CMS collaborations.

In May 2020, available cloud resources of JINR and its Member States were included in the project. More than 5,800 work units were performed on the clouds in two months, half of which were carried out on the JINR cloud, and the other half were carried out on the clouds of the following organizations:

- Plekhanov Russian University of Economics,
- Research Institute for Nuclear Problems of Belarusian State University,
- Khetagurov North Ossetian State University,
- Institute of Physics of the Azerbaijan National Academy of Sciences,
- Institute for Nuclear Research and Nuclear Energy of the Bulgarian Academy of Sciences.

Understanding the structure of SARS-CoV-2 proteins is essential, since it can accelerate the discovery of new therapeutic methods of treatment for COVID-19. The study of the structure of proteins encompasses methods of molecular dynamics. They enable the investigation of the whole range of states that a protein takes; however, the implementation of these methods requires significant computing resources. Modern datasets generally comprise several microseconds of the simulation of a single protein. Some sets last milliseconds. Nevertheless, many processes in proteins occur over longer periods of time. Thus, the simulation of all SARS-CoV-2 related proteins with a duration sufficient for their biological analysis demands an unprecedented amount of computing resources.

To solve this problem, more than a million people and organizations worldwide volunteer computing resources of their computers to perform simulations of SARS-CoV-2 proteins. The work of such a collaboration became possible due to the Folding@Home distributed computing platform, which has broken the Exaflop barrier in 2020 to turn into the largest distributed supercomputer in the world. Using this resource, datasets with a duration of tenths of seconds were obtained. These data confirmed by various experimental observations are already available at https://covid.molssi.org/.

Folding@Home is a community of volunteers, researchers and organizations that help with their intelligent and computing resources to understand the dynamics of proteins, their functions and dysfunctions in order to find new proteins and drugs. The project was founded in 2000 to study protein folding. Back then, the simulation of the folding of even small proteins on one computer could easily take thousands of years. To handle that challenge, Folding@Home divided one huge problem into a great number of small simulations, i.e. work units, which can be carried out independently of each other.

Over the years, the project evolved and expanded the range of solved problems related to cancer, resistance to microbes and diseases associated with the dysfunction of protein membranes. When SARS-CoV-2 appeared, Folding@Home changed its focus. Many people and organizations around the world found a way to act despite the circumstances in which one might feel helpless. In less than three months, the number of active computing devices in the system has grown from 30,000 to one million. Not only volunteers, but also organizations joined the project; NVIDIA, Amazon, CERN, Intel and many others were able to allocate available resources for Folding@Home tasks.

JINR and its Member States decided to make their contribution. Initially, it was done within their participation in the processing of data from the Large Hadron Collider. The major idea lied in using available computing resources of the grid. With the idea of using available resources, Folding@Home work units were launched on the clouds of JINR and its Member States. After two months of work, 5,800 work units were performed on these clouds.

The priorities during the launch of these work units encompassed the following: elimination of the influence on virtual machines already running on the clouds; ability to quickly remove the Folding@Home load; acquisition of data on the amount of work performed on each of the cloud infrastructures; top priority to the tasks of studying SARS-CoV-2.

The resources of the distributed cloud infrastructure of JINR and its Member States are combined into a single environment via the DIRAC Interware. The use of this platform allowed us to significantly reduce the work related to work unit management and gave complete control over their starting and stopping.

For accounting the allocated resources, a team "Joint Institute for Nuclear Research" was created in Folding@Home; it combines separated accounts of each of the participants of the JINR distributed cloud infrastructure. The resource consumed by a specific work unit is counted as a contribution of the corresponding cloud infrastructure and the JINR team. Owing to it, on the Folding@Home statistics page, one can find out the exact number of work units performed on different clouds, as well as the total number of work units carried out by the whole team.

After two months of working on the clouds of JINR and its Member States, about 6 thousand work units were performed, and the overall workload in CPU hours even exceeded twice the JINR contribution under the ALICE experiment. Folding@Home work units, launched on the clouds, revealed the actual performance of different cloud infrastructures. This made it possible to detect bottlenecks of the infrastructure and eliminate them.

Work performed by all clouds integrated into DIRAC:
https://stats.foldingathome.org/team/265602

Work performed by CERN:
https://stats.foldingathome.org/team/38188

Contribution of JINR together with the ALICE experiment:
https://stats.foldingathome.org/donor/84204021

The present article relies on the article "Citizen Scientists Create an Exascale Computer to Combat COVID-19":
https://www.biorxiv.org/content/10.1101/2020.06.27.175430v1.full.pdf.