

SIMULATION OF JOB EXECUTION IN DISTRIBUTED HETEROGENEOUS COMPUTING INFRASTRUCTURES

*I. S. Pelevanyuk*¹, *D. Campis*²

Joint Institute for Nuclear Research, Dubna

Execution of one computing job demonstrates that software is correctly working. But, when the same job has to be executed thousands of times, it may cause different issues. Nowadays special distributed heterogeneous computing infrastructures are widely used for this type of workload. The main issue when running big workloads on them is network limits. These limits may be imposed in different levels: server, cluster, and storage level. With limited network there is a threshold after which increment of CPU resources does not speed up jobs execution rate. The purpose of this work is creation of a software platform for simulation of job execution in distributed computing infrastructures. The purpose of the system is a prediction of job execution rate in real infrastructure. The software platform was developed and tested. Python programming language was used for development. InfluxDB is used for results storage and visualization.

Выполнение одной вычислительной задачи демонстрирует корректность работы программного обеспечения. Но когда та же самая задача должна быть выполнена в тысячах экземпляров, это может вызывать ошибки. Для такого массивного запуска активно используются распределенные гетерогенные вычислительные инфраструктуры. Основной вопрос при работе большого количества однотипных задач — это ограничения, связанные со скоростью сети. Эти ограничения могут возникать на разных уровнях: сервера, кластера или системы хранения. С ограниченной пропускной способностью сети существует порог, после которого добавление дополнительных вычислительных ресурсов не ускорит выполнение пакета задач. Цель данной работы — создание программной платформы для симуляции выполнения задач в распределенной гетерогенной вычислительной сети. Основная задача разрабатываемой системы — прогноз скорости выполнения задач на доступных вычислительных инфраструктурах. Система разработана и протестирована. В качестве языка программирования использовался Python. База данных InfluxDB используется для хранения и последующей визуализации полученных данных.

PACS: 44.25.+f; 44.90.+c

Received on November 14, 2022.

¹E-mail: pelevanyuk@jinr.ru

²E-mail: daniel.campis100@gmail.com