

## MIDDLEWARE FOR BIG DATA PROCESSING: TEST RESULTS

*I. Gankevich*<sup>1</sup>, *V. Gaiduchok*, *V. Korkhov*, *A. Degtyarev*, *A. Bogdanov*

St. Petersburg State University, St. Petersburg, Russia

Dealing with large volumes of data is resource-consuming work which is more and more often delegated not only to a single computer but also to a whole distributed computing system at once. As the number of computers in a distributed system increases, the amount of effort put into effective management of the system grows. When the system reaches some critical size, much effort should be put into improving its fault tolerance. It is difficult to estimate when some particular distributed system needs such facilities for a given workload, so instead they should be implemented in a middleware which works efficiently with a distributed system of any size. It is also difficult to estimate whether a volume of data is large or not, so the middleware should also work with data of any volume. In other words, the purpose of the middleware is to provide facilities that adapt distributed computing system for a given workload. In this paper we introduce such middleware appliance. Tests show that this middleware is well suited for typical HPC and big data workloads, and its performance is comparable with well-known alternatives.

Работа с большими объемами данных требует значительных ресурсов и все чаще и чаще делегируется не одному компьютеру, а целой распределенной вычислительной системе. По мере увеличения числа компьютеров в распределенной системе увеличиваются и усилия для эффективного управления системой. Когда система достигает некоторого критического размера, значительные усилия должны быть направлены на улучшение ее отказоустойчивости. Трудно оценить, в какой момент распределенная система нуждается в таких возможностях для имеющейся нагрузки, поэтому механизмы отказоустойчивости должны быть реализованы на уровне промежуточного программного обеспечения, которое эффективно работает с распределенной системой любого размера. Кроме того, трудно оценить, является объем данных большим или нет, так что промежуточное программное обеспечение должно также работать с данными любого объема. Другими словами, целью промежуточного программного обеспечения является предоставление средств, которые адаптируют распределенную вычислительную систему к имеющейся рабочей нагрузке. В данной работе мы представляем такое промежуточное программное обеспечение. Тесты показывают, что это программное обеспечение промежуточного слоя хорошо подходит для типичных нагрузок приложений высокопроизводительных вычислений и больших данных, а его производительность сравнима с известными аналогами.

PACS: 07.05.-t

Received on December 9, 2016.

---

<sup>1</sup>E-mail: i.gankevich@spbu.ru