

# 10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)



Contribution ID: 344

Type: **not specified**

## Высокоскоростной Data-Lake для мега-сайенс проекта NICA

*Tuesday, 4 July 2023 17:15 (15 minutes)*

Мега-сайенс проект NICA задаёт высокую планку к вычислительным ресурсам и системам хранения и обработки данных. Участники коллабораций MPD, VM@N, SPD при выполнении расчётов активно задействуют различные вычислительные ресурсы ОИЯИ: МИВК Tier-2, СК Говорун, Облако JINR-Cloud, вычислительный кластер NCX. При выполнении расчётов применяется классическая иерархия систем хранения и обработки данных. В качестве "холодной" системы для хранения данных используется распределённая файловая система EOS, в качестве "теплой" и "горячей" систем для обработки данных — файловые системы NFS/ZFS и Lustre, привязанные к вычислительным ресурсам. Поиск свободных вычислительных ресурсов требует от участников коллабораций постоянно выполнять перенос своих "горячих" данных с одного вычислительного ресурса на другой.

В рамках данной работы предложено решение на основе файловой системы Lustre для быстрого копирования "горячих" данных между СК Говорун и вычислительным кластером NCX. Разработанная архитектура включает в себя следующие сегменты для обработки данных: локальная файловая система для СК Говорун, локальная файловая система для вычислительного кластера NCX, а также сегмент с зеркальной файловой системой, доступной на обоих вычислительных ресурсах. Данное решение позволит существенно сократить усилия пользователей и время на перенос "горячих" данных. Также в работе рассмотрены режимы отказоустойчивости на основе компонент файловой системы Lustre, сервисов Heartbeat и DRBD.

### Summary

**Primary authors:** KOKOREV, Aleksander; DOLBILOV, Andrey (JINR); BELYAKOV, Dmitry (JINR); PODGAINY, Dmitry (JINR); SLEPOV, Ivan (JINR); LUBIMOVA, Maria; MATVEYEV, Mikhail (JINR)

**Presenter:** BELYAKOV, Dmitry (JINR)

**Session Classification:** Computing for MegaScience Projects

**Track Classification:** Computing for MegaScience Projects