

Бочков Максим Константинович

студент

Андрианов Емельян Ахмедович

студент

Зайнуллина Татьяна Сергеевна

студентка

ФГАОУ ВПО «Национальный исследовательский

ядерный университет «МИФИ»

г. Москва

ИСПОЛЬЗОВАНИЕ РАСПРЕДЕЛЕННЫХ ФАЙЛОВЫХ СИСТЕМ В КАЧЕСТВЕ СЕТЕВЫХ ХРАНИЛИЩ ДАННЫХ

***Аннотация:** в данной работе анализируются распределенные файловые системы в качестве сетевых хранилищ данных. В статье рассматриваются основные свойства распределенных файловых систем.*

***Ключевые слова:** распределенная файловая система, отказоустойчивость, масштабируемость, прозрачность, hdfs, сетевое хранилище данных.*

В настоящее время одной из самых актуальных проблем оптимизации затрат на хранение информации в крупных компаниях является высокая стоимость дискового пространства. Однако есть решение. Распределенные файловые системы обеспечивают надежное хранение данных при невысокой стоимости оборудования, причем особую эффективность эти системы приобретают при работе на крупных кластерах. Именно поэтому крупные компании заменяют старые системы хранения данных на распределенные файловые системы. В перспективе, такая модернизация принесет финансовую выгоду.

Изучение каждой распределенной файловой системы в отдельности трудоемкая задача. Но ответить на вопрос, какую из них предпочтительней использовать для модернизации, возможно только исследуя достоинства и недостатки нескольких распределенных файловых систем и проводя сравнительный анализ.

Все распределенные файловые системы обладают тремя очень важными свойствами: масштабируемость, прозрачность и защищённость от сбоев. Данные свойства дают возможность организовать надежное хранилище данных.

В настоящее время существует порядка тридцати распределенных файловых систем. Для участия в исследования были отобраны распределенные файловые системы, по следующим критериям:

- отсутствие платы за использование программного обеспечения распределенной файловой системы;
- наличие активного, развивающегося сообщества;
- наличие репликации блоков, которая обеспечивает программную отказоустойчивость;
- наличие документации на английском языке;
- поддержка FUSE или NFS доступа;
- наличие релиза файловой системы;
- наличие алгоритма обработки сбоев.

Всем критериям по результатам сравнения удовлетворили наиболее перспективные: HDFS и Ceph с централизованной архитектурой, MooseFS и GlusterFS с децентрализованной архитектурой.

В качестве методики тестирования масштабируемости измерены скорости записи в зависимости от количества серверов, на которых развернуты распределенные файловые системы, а также количества клиентов, одновременно производящих запись. Для тестирования файловой системы на отказоустойчивость отключались сервера, входящие в состав кластера, и проведено слежение за балансировкой и за сохранностью данных.

Выбранные файловые системы обеспечивают масштабируемость и устойчивость при сбоях с помощью различных методов, но с одинаковой эффективностью. Основное различие между ними заключается в их архитектуре. Децентрализованные файловые системы работают лучше, чем централизованные, в силу распределения управления загрузкой серверов. Более того, выбор той или иной распределенной файловой системы в первую очередь зависит от того, для каких

целей она будет использоваться. Для более высокой производительности предпочтительными являются асинхронная репликация и использование индексов для структурирования пространства имен файловых систем HDFS и MooseFS. В свою очередь, для управления большим количеством данных и запросов к ним более эффективной является децентрализованная архитектура Ceph и GlusterFS.

Список литературы

1. Nicolae B., Antoniu G., Boug'e L., Moise D., Carpen-amarie R. Blobseer: Nextgeneration data management for large scale infrastructures. J. Parallel Distrib. Comput. – 2011.
2. Shvachko K., Kuang H., Radia S., Chansler R. The hadoop distributed file system. In: Proceedings of the 2010 IEEE 26th Symp. On Mass Storage Systems and Technologies (MSST), Washington, DC, USA, IEEE Computer Society.
3. Redundant Arrays of Inexpensive Disks 2014, ARPACI-DUSSEAU.