

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ СЖАТИЯ/УПЛОТНЕНИЯ ДАННЫХ В ОБЛАЧНЫХ СИСТЕМАХ ХРАНЕНИЯ ДАННЫХ НА ПЛАТФОРМЕ CEPH

Роман А.Б.

(Санкт-Петербург, СПбГ НИУ ИТМО)

Научный руководитель – к.ф.-м.н., декан ф-та ИКТ, доцент кафедры СиОТ

Сергей Э.Х.

(Санкт-Петербург, СПбГ НИУ ИТМО)

На сегодняшний день объём данных, хранящихся у среднестатистического пользователя, продолжает неуклонно расти. Это обусловлено большим количеством факторов, к числу которых можно отнести снижение цен на дисковое пространство, а также постоянное увеличение объёма, занимаемого создаваемым и потребляемым контентом. Вместе с этим растёт и потребность в сохранности информации, однако создание, хранение и поддержание актуальности резервных копий становится всё более ресурсоёмкой задачей. Кроме того, усиливается и потребность в быстром доступе к данным из любой точки в любой момент времени.

В результате этого, огромную популярность приобрела так называемая концепция облачного хранения данных. В рамках данной модели пользователю предлагается хранить свою информацию на удалённых серверах, предоставляемых некоторым третьим лицом, выступающим в роли поставщика. Таким образом, хранение информации становится доступным конечному пользователю в качестве сервиса, что является отражением так называемой бизнес-модели STaaS (Storage as a service), в рамках которой крупные компании, обладающие огромными хранилищами данных, предоставляют в аренду пользователям простаивающее дисковое пространство по принципу подписки.

Очевидно, что для поставщиков облачных сервисов ключевыми являются проблемы, связанные с распределённым хранением данных. Существует высокая потребность в высокопроизводительных и надёжных системах хранения данных, с высокой производительностью файловой системы, способной обеспечить параллельный доступ к файлам более чем для тысяч клиентов.

Одной из таких систем является Ceph. Ceph использует несколько интересных решений в своем архитектурном подходе. Разработчики Ceph пытались создать систему хранения данных, удовлетворяющую следующим ключевым условиям:

- Легкость масштабирования до петабайтных размеров;
- Высокая производительность при различных нагрузках (числа операций ввода/вывода в секунду и различной полосе пропускания);
- Повышенная надёжность;
- Высокая доступность данных, то есть минимальное время простоя при авариях.

Ceph предоставляет на выбор три различных абстракции для работы с хранилищем:

- Абстракцию объектного хранилища (RADOS Gateway);
- Абстракцию блочного устройства (RADOS Block Device);
- Абстракцию POSIX-совместимой файловой системы (CephFS).

В своей работе я рассмотрел Ceph, как абстракцию блочного устройства (в оригинале - RADOS Block Device, или RBD), которая предоставляет пользователю возможность создавать и использовать виртуальные блочные устройства произвольного размера. Программный интерфейс RBD позволяет работать с этими устройствами в режиме чтения/записи и выполнять служебные операции - изменение размера, клонирование, создание и возврат к снимку состояния и т.д.

Также, в связи с ростом объёмов передаваемых данных, постоянно будет востребована и возможность сжатия данных, с целью повышения производительности и минимизирования нагрузки на сеть. Ceph также идеально подходит в данном случае, т.к. имеет встроенную возможность сжатия, а именно BlueStore. BlueStore поддерживает

встроенное сжатие с использованием snappy, zlib или lz4. Сжатие данных в BlueStore определяется комбинацией режима сжатия и любых подсказок, связанных с операцией записи.

Таким образом целью данной работы является исследование возможностей сжатия данных, непосредственно в облачных системах хранения данных, с использованием возможностей платформы Ceph, а именно, с использованием хранилища объектов BlueStore.

В ходе работы были изучены особенности современных облачных систем хранения данных, подробно была разобрана архитектура распределённой файловой системы Ceph и её возможности, используемые для сжатия данных.

Заключение. В результате проделанной работы была развёрнута и настроена система Ceph, с применением хранилища объектов BlueStore и были проведены исследования возможностей сжатия данных.

Автор

_____ / Р.А. Балдук /

Научный руководитель

_____ / С.Э. Хоружников /

Заведующий кафедрой СиОТ

_____ / С.Э. Хоружников /