СХД для потоковых данных: проблемы и решения

Выбирая систему хранения для приложений, использующих преимущественно потоковые данные, приходится учитывать большое количество характеристик. Но нередко при выборе специализированных СХД под такие задачи бренды вместе с избыточностью параметров навязывают и высокую стоимость решения. Выбор и приоритезация критериев оценки СХД для работы с потоковыми данными – тема публикации.

Полина Трофимова – директор по маркетингу и продажам, компания AVRORAID.

Введение

Традиционно принято считать, что при выборе СХД необходимо достичь оптимального соотношения производительности, доступности (надежного хранения и отказоустойчивого доступа) и итоговой стоимости хранения.

Высокая производительность – это необходимое условие эффективной работы приложений. Производительность СХД представлена двумя показателями скорости: операции чтения/записи, то есть количеством операций ввода/вывода в секунду (IOPS) и суммарной скоростью передачи данных (Мбайт/с).

Надежное хранение и отказоустойчивый доступ следует также всегда принимать во внимание, поскольку именно эти параметры гарантируют непрерывность бизнес-процессов. Как правило, надежность повышается благодаря использованию дорогостоящих аппаратных компонент с высоким сроком наработки на отказ (сроком службы). Доступность здесь – главное требование: данные должны быть доступны всегда. Отсутствие доступа к данным в большинстве случаев равноценно потере данных. Доступ к данным может отсутствовать как в случае отказа технических средств, выхода из строя канала, так и в случае отсутствия необходимой производительности для выполнения прикладных задач. Обеспечить доступность можно за счет дублирования аппаратных компонент, которое приводит к удорожанию решения, а также за счет применения различных уровней RAID.

Стоимость хранения – это отношение суммарной стоимости системы, иногда включающее и стоимость обслуживания, к предоставляемому объему полезного дискового пространства (руб./Мбайт). Обычно организации стараются минимизировать стоимость хранения за счет уменьшения начальной стоимости приобретения СХД: использование недорогих, но при этом медленных дисков, использование экономичного RAID 5, который позволяет выдерживать отказ только одного диска и т.д.

Как достичь оптимального соотношения между перечисленными выше параметрами?

Классификация СХД по типу обрабатываемых данных

С точки зрения решаемых СХД задач, все приложения условно подразделяются на:

– приложения со случайным доступом к данным (транзакционные системы);
– приложения с потоковым доступом к данным (потоковые приложения).

Для транзакционных систем со случайным доступом характерно большое количество запросов на операции чтения и записи небольшими блоками данных, например, база данных, имеющая размер блока 8-32 Кбайт. Наиболее действенным механизмом увеличения производительности для транзакционных приложений является кэширование. Так как передается небольшой объем данных, кэширование "объединяет" маленькие блоки в большие порции данных, которые дисковая подсистема может записать за один раз. Таким образом, кэширование сокращает количество обращений к дисковой подсистеме самому медленному компоненту любой СХД. Важной составляющей также является возможность настройки размера страйпа. Изменение значения этой переменной существенно улучшает характеристики производительности СХД.

Потоковый доступ – это последовательная запись или чтение блоков данных большого размера. Для потоковых приложений характерны запросы на операции с большими блоками данных (от 512 Кбайт и более). Ввиду большого размера блоков и последовательности обработки данных, кэширование не дает существенного выигрыша в производительности операций записи/чтения. Очевидно, что для приложений с потоковым доступом самым узким местом является RAID-контроллер, выполняющий расчет контрольных сумм.

Характеристиками производительности СХД являются IOPS (количество операций ввода-вывода в секунду) и Мбайт/с (объем переданных данных в секунду). Показатель IOPS приводится обычно для операций с блоками небольшого размера. Значение показателя Мбайт/с указывается для нагрузки крупными блоками.

Для примера, на рис. 1 представлены результаты операции последовательного чтения RAID0 (один из самых быстрых для последовательного чтения/записи), 2 SSD-диска. Соответственно, по горизонтали –

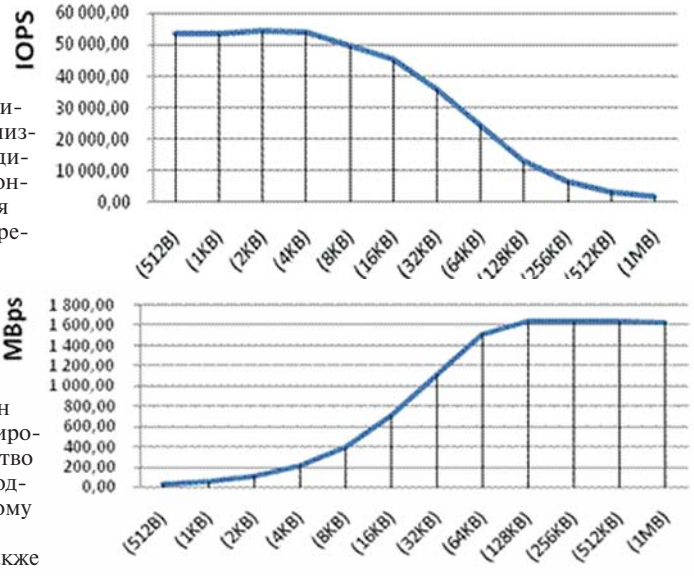

Рис. 1. Последовательное чтение, RAID 0, 2 SSD диска. Зависимость количества операций в секунду и скорости передачи от размеров блока данных.

размер блока, по вертикали – IOPS, или мегабайты в секунду. Блок меньше 128 Кбайт можно соотнести с нагрузкой БД или веб-сервера, блок большего объема – с работой с файлами (потоковые данные). Диаграммы отражают характерную зависимость количества операций в секунду (IOPS) и скорости передачи от размеров блока данных. Видно, что при увеличении размера блока количество IOPS уменьшается, так как возросший объем передаваемых данных (Мбайт/с) создает нагрузку на вычислительные мощности СХД.

Логически можно предположить, что необходимая скорость потокового доступа должна обеспечиваться высокой скоростью интерфейса и мощностью RAID-контроллера (см. рис. 1).

Какие параметры действительно необходимо учитывать при выборе СХД для работы с потоковыми данными? Рассмотрим требования в разрезе решаемых задач.

Особенности приложений, работающих с потоковыми данными

Круг задач, в которых нагрузка на систему хранения обеспечивается потоковыми данными, довольно широк: резервное копирование, нелинейный видеомонтаж, видеонаблюдение, системы документооборота.

Резервное копирование и восстановление

Стремительно возрастающие объемы данных, измеряемые десятками терабайт, с одной стороны, неуклонно увеличивают "технологическое окно", необходимое для выполнения операций копирования. С другой стороны, размер окна сокращается в режиме работы компании "24x7". Отсутствие в инфраструктуре узких мест на уровне работы дисковой подсистемы и каналов передачи данных гарантирует минимальное время не только выполнения резервной копии, но и ожидаемое время восстановления данных из backup'а. Время инициализации и реконструкции дискового массива в задачах резервного копирования также является серьезным ограничением, которое имеет смысл принимать во внимание.

Обработка видео (нелинейный монтаж – post production)

Нелинейный монтаж подразумевает работу с большими объемами аудио- и видеоматериалов в цифровом формате. Основными особенностями этой задачи являются работа с потоковыми данными крупными блоками, требования к обеспечению высокой производительности обработки данных, чувствительность к изменению скорости обмена данными с СХД, значительные объемы данных на один LUN, длительное надежное хранение данных в условиях ограничения бюджета.

При выборе СХД для задач нелинейного монтажа, также как и под резервное копирование, следует ориентироваться на **показатель скорости потокового чтения/записи**. Соответственно, чем выше скорость потокового чтения/записи на фоне гарантированной степени надежности и обеспечиваемого объема хранилищ, тем эффективнее будет выполняться нелинейный монтаж. Как правило, задачи post-production выполняются в условиях ограничения времени и четких сроков передачи результата. А диски, как известно, всегда выходят из строя в самый неподходящий момент. Поэтому немаловажным параметром для СХД под задачи обработки видео является **время реконструкции массива**. СХД должна обеспечивать быструю реконструкцию. Для большинства аппаратных СХД время реконструкции составляет 10-20 часов, что является серьезным ограничением в условиях монтажа. Здесь неоспоримое преимущество имеют программные СХД. Так, в решениях компании AvroRAID, благодаря уникально реализованному алгоритму расчета RAID6, скорость реконструкции составляет всего **3,5 часа**!

Возможность установки приоритета приложениям и сервисным функциям также позволяет сохранять заявленный уровень производительности и не снижать скорость обмена данными при выходе из строя 1 или 2 дисков. Эта возможность очень востребована именно для задач нелинейного монтажа, так как "проседание" производительности отрицательно сказывается на качестве работ. Для сохранения показателей производительности также выполнена функция Advanced Recontruction, которая описана ниже.

Тестирование

Специалистами компании AvroRAID было проведено тестирование типичной модели гетерогенного ИТ-парка небольшой студии по созданию и обработке медиаконтента. Задачи тестирования:

- демонстрация высоких показателей производительности системы хранения на основе ПО AVRORA при работе с типичной нагрузкой для видеомонтажа с применением ПО MetaSAN для предоставления функции разделяемого доступа;
- подтверждение стабильности работы СХД (неизменность значений скорости чтения и записи) при выходе дисков из строя и при выявлении системой медленно (некорректно) работающего диска;
- замер времени восстановления (реконструкции) массива.

Описание стенда

Три рабочих станции под управлением Windows Server 2008 R2 с установленным ПО MetaSAN версии 4.6.0 подключены напрямую по двум каналам FC 8Gb к СХД с 24 дисками SATA (рис. 2).

Нагрузка обеспечивалась стандартным средством моделирования нагрузки – IOMeter – с настройкой различных паттернов.

Описание аппаратной части и настроек тестирования СХД на основе ПО AVRORA даны в табл. 1, характеристики рабочих станций – в табл. 2, сценарии и результаты испытаний – в табл. 3.

Результаты тестирования

Система хранения данных на основе ПО AVRORA с 24 дисками SATA при прямом подключении по FC 8 Гбайт/с и нагрузке

Рис. 2. Конфигурация стенда при проведении тестирования СХД на основе ПО AVRORA.

с трех рабочих станций показала значительное превосходство в части характеристик производительности в сценариях работы с медиаконтентом:

- производительность СХД на основе ПО AVRORA составила 2,6–3,2 Гбайт/с на чтение и запись потоковых данных в RAID6;
- потери производительности при выходе из строя диска отсутствуют;

Табл. 1. Описание СХД на основе ПО AVRORA

Характеристика	Значение
Шасси	Шасси SuperMicro
Внешний интерфейс	FC 8Gb/s
Количество портов	6 портов
Количество дисков	24 диска
Тип дисков	SATA 7200 1.5ТБ
Размер кэш	10ГБ
Тип RAID	RAID6 на все диски, один LUN
Размер LUN	33ТБ
Тип LUN	LUN отформатирован NTFS
Параметры MPIO	MPIO, режим «минимальная глубина очереди»
Версия ПО AVRORA	2280_1.2.50

Табл. 2. Описание рабочих станций

Характеристика	Значение
Сервер	Supermicro Superserver 6014H
ОС	Windows 2008 Server
Интерфейс подключения к СХД	FC 8Gb/s
Количество портов подключения к СХД	2 порта

Табл. 3. Результаты испытаний

Сценарий испытаний	Результат	Оценка
Запись в 3 файла размером 50 ГБ	1400 МБ/сек 340 МБ/сек 640 МБ/сек	Суммарная производительность 2.4ГБ/сек
Параллельное чтение и запись с трех файлов размером 80 ГБ с 3 рабочих станций по 2 потока на чтение и 1 поток на запись	860 МБ/сек 950 МБ/сек 870 МБ/сек	Суммарная производительность 2.6 ГБ/сек
Параллельное чтение из одного файла размером 150ГБ с 3 рабочих станций по 2 каналам	1150 МБ/сек 820 МБ/сек 990 МБ/сек	Суммарная производительность 2.96 ГБ/сек
Включение режима Reconstruct in Advance of Drive Completion при работе системы со сбойным диском	Из-за низкой производительности сбойного диска произво-дительность чтения с СХД не превышала 1,5ГБ/сек. При включенном механизме Reconstruct in Advance of Drive Completion производительность чтения повысилась до 3 ГБ/сек.	Режим Reconstruct in Advance of Drive Completion гарантирует постоянные значения производительности даже при наличии сбойных дисков в массиве.
Сохранение показателей производительности при выходе из строя 2 дисков	При отключении 2 дисков суммарная производительность на чтение и запись осталась прежней (Около 2.6 ГБ/сек)	Производительность системы не изменяется при выходе из строя 2 дисков
Сохранение показателей производительности при запуске реконструкции	При начале реконструкции производительность осталась прежней.	При запуске реконструкции производительность не изменяется
Выполнение реконструкции массива	При отсутствии постоянной нагрузки восстановление массива завершилось за 3,6 часа	Время реконструкции массива из 24 дисков – менее 4 часов

– восстановление (реконструкция) массива выполняется менее чем за 4 часа (что в 6 раз быстрее, чем у конкурентных решений в этом же ценовом сегменте).

Высокая надежность большого объема хранения данных – еще одно важное требование СХД для работы с видеоконтентом. Самым экономичным, с точки зрения использования дополнительного (дублирующего) дискового пространства, является уровень RAID5. Технология RAID5 задействует небольшой объем вычислительных мощностей процессора (рассчитывается одна контрольная сумма), поэтому обеспечивает скорости, достаточные для эффективной работы с потоковыми данными. Однако по степени отказоустойчивости RAID5 значительно уступает RAID6. RAID5 не спасает в случае последовательных отказов дисков, а выдержать может отказ лишь одного диска.

Принципиальное отличие технологий объединения дисков в единое пространство – RAID6 и RAID5 – заключается в том, что в RAID6 вычисляются две контрольные суммы, а не одна. Это позволяет в массиве RAID6 восстановить данные после выхода из строя двух жестких дисков.

Недостаток RAID6 – низкая производительность. Действительно, двойная схема четности, реализованная в RAID6, требует больших вычислительных мощностей от контроллера. Как следствие, проявляется снижение общей производительности системы. Разумеется, на одной и той же аппаратной платформе RAID6 будет работать медленнее, чем RAID5. Критичность данного факта ставится под сомнение в случае, если производительности RAID-контроллера достаточно для обслуживания имеющегося массива дисков, и он не является узким местом в системе. Высокопроизводительный RAID-контроллер также способствует увеличению стоимости системы. В качестве альтернативы для достижения оптимального соотношения стоимости, надежности, производительности могут выступать программные RAID. СХД, использующие программные RAID, позволяют эффективно использовать технологию RAID6, сводить к минимуму проявление ее недостатков: низкая производительность и высокая стоимость хранения. Так, в решениях компании AvroRAID уникально реализованный алгоритм расчета RAID6 позволяет получать рекордные показатели скоростей на потоковых нагрузках (**4 Гбайт/с**). В дополнение к высокой производительности в программных RAID6 от AvroRAID можно отметить следующие полезные функции:

– *Advanced Reconstruction* – функция, сохраняющая производительность RAID6 за счет автоматического исключения из операции чтения диска, производительность которого по каким-либо причинам упала ниже допустимой. Данные на диске остаются консистентными. При этом производительности RAID6 достаточно для того, чтобы с помощью контрольных сумм восстановить информацию, хранящуюся на медленном диске, нежели считывать ее;
– *высокая скорость реконструкции массива* – около 3 часов на RAID6 (по данным AvroRAID) в сравнении с 10-20 часами в аппаратных СХД на RAID6;
– *отсутствие снижения производительности СХД в момент реконструкции* – отсутствие задержек в производственных процессах.

Электронный документооборот

Помимо требований к обслуживанию ядра системы документооборота, которое обычно соответствует типу транзакционной системы (высокие показатели IOPS на случайное чтение небольшими блоками), во многих продуктах документооборота существует также возможность вынести для хранения на отдельном устройстве крупных документов (отсканированных копий многостраничных договоров, графические документы и другие крупные файлы, прикрепляемые к карточкам).

Например, в продукте DocsVision, использующем в качестве СУБД MS SQLServer 2008, хранение крупных документов может быть организовано на отдельной СХД, оптимизированной для работы с потоковыми данными. Положительный опыт использования был подтвержден в процессе нагрузочного тестирования системы DocsVision для Банка ВТБ24, которое было проведено компаниями Syntellect, премиум-партнер компании "ДоксВижн", и Microsoft. Нагрузочный стенд был развернут на базе инфраструктуры Технологического центра Microsoft в Москве (МТС).

При проведении тестирования была создана нагрузка, эмулирующая одновременную работу с системой **5 000** пользователей. В соответствии со сценариями тестирования пользователи выполняли более 10 различных типов операций, включающих как работу с карточками и документами, так и операции сканирования, обмена данными с внешними приложениями, а также поиск по всему массиву данных. Интенсивность операций составляла более **2 млн в час**.

Основные параметры стенда (количество карточек документов в базе данных):

– 60 млн (220 млн записей в БД);
– 50 млн сканированных изображений различного размера;
– 5 Тбайт (физический объем БД).

Вместо заключения

Несколько полезных советов при выборе СХД для потоковых данных:

– *большой объем доступной кэш-памяти слабо влияет на производительность операций потокового чтения/записи;*
– *рекордное количество IOPS на маленьких блоках не гарантирует скоростей потокового доступа на крупных блоках;*
– *производительность СХД обычно заявляется на наиболее простых уровнях RAID;*
– *допустимое время реконструкции массива и отсутствие "проседания" при выходе дисков из строя – обычно не заявляются производителем и требуют уточнения;*
– *для работы с видео необходимо убедиться в достаточности заявленного максимального размера создаваемого LUN.*

Полина Трофимова,
директор по маркетингу и продажам,
компания AVRORAID