

4. Хранение мультимедийных данных

4.1. Организация хранения

4.2. Магнитные запоминающие устройства

4.3. Оптические запоминающие устройства

4.4. Файловые системы для оптических носителей

4.5. Иерархическое управление запоминающими устройствами

Организация хранения

Общие требования:

- Большая емкость (вместительность)
- Высокая скорость передачи данных

Магнитные диски с прямым доступом:

- Скоростные, но потенциально недостаточная емкость

Сменные оптические накопители (дисководы с автоматической сменой дисков):

- Огромная емкость, но низкая скорость передачи (особенно запись)

Заключение:

- Необходим 'тесный союз' дискового и оптического способов хранения
- В ММСУБД должно быть интегрировано несколько структур хранения данных
- Ожидается, что ММСУБД работает на нескольких серверах и использует несколько устройств управления памятью

Организация хранения

Организация хранения мультимедийных данных:

1. Обычная СУБД (реляционная/объектно-ориентированная) на магнитных дисках:
 - Структурированные данные
 - Подойдет для небольших ММСУБД
2. Обычная СУБД с высокопроизводительной параллельной дисковой системой:
 - Оптимизированная производительность
 - Подойдет для ММСУБД средних размеров
3. СУБД с оптической системой хранения:
 - Полуоперативное хранение данных размером в несколько терабайт
 - С учетом автономного хранения – нет пределов в размере данных
4. СУБД, сопряженная с медиа-сервером (хранение на магнитных и оптических носителях):
 - Сложности с одновременным доступом к подсистемам хранения
5. Распределенная СУБД с независимыми подсистемами хранения:
 - Поддержка целостности бд

Магнитные запоминающие устройства

Достоинства:

- Всегда доступна
- Неограниченное число перезаписей
- Лучшая производительность (на данный момент)

Проблемы:

- Надежность падает со временем: решение - избыточность
- Пропускная способность необходимая для мультимедиа может превосходить возможности обычных магнитных дисков: решение – распараллеливание

Магнитные запоминающие устройства

Технология RAID (**R**edundant **A**rray of **I**nexpensive/**I**ndependent **D**rives) - избыточный массив недорогих/независимых дисков¹:

- Группа небольших дисков функционирует быстрее и надежнее в сравнении с одним большим жестким диском
- Необходима серверам, предоставляющим широкополосные сервисы (например, видео по требованию)
- Может быть сопряжена с более медленным оптическим хранением (большой емкости)
- Для распараллеливания диски должны иметь свои собственные контроллеры

¹ - <http://www.ixbt.com/storage/raids.html>

Магнитные запоминающие устройства

Семь типов RAID массивов:

RAID 0:

- Распределение данных по различным дискам массива
- Запросы на чтение/удаление транслируются в несколько подзапросов, выполняемых различными дисками
- Наилучшая производительность среди все RAID типов
- Недостаток: отсутствует отказоустойчивость (выход из строя одного диска ведет к невозможности работать со всем массивом)
- При сопряжении со сменными оптическими накопителями отказоустойчивость становится менее важной

RAID 1:

- Зеркализация дисков (дублирование данных на два диска, рассматриваемых системой как один): повышается надежность
- Операции записи производятся одновременно в оба диска
- Наиболее дорогой из RAID типов
- Возможно параллельное чтение с двух дисков

Магнитные запоминающие устройства

RAID 2:

- Чередование данных по группе дисков, часть из которых используется для хранения контрольной информации (обнаружение и коррекция ошибок)
- Например, 10 дисков с данными плюс 4 диска с контрольной информацией (диски четности)
- Медленное восстановление
- Высокая скорость передачи данных больших объемов
- Только одна команда ввода/вывода к дисковому массиву за раз (но быстрое выполнение)
- Низкая скорость обработки запросов (не для систем ориентированных на обработку транзакций)

Магнитные запоминающие устройства

RAID 3:

- Чередование данных по группе дисков с одним диском для контрольной информации
- Большинство контроллеров диска могут определить место сбоя (поэтому количество дисков с контрольной информацией, требуемых RAID 2, можно уменьшить до одного)
- В случае отказа на одном из дисков данные могут быть восстановлены путем XORа данных на остальных дисках
- Отказ диска мало влияет на скорость работы массива
- Медленное восстановление

Магнитные запоминающие устройства

RAID 4:

- Расслоение (чередование) данных выполняется на уровне секторов, а не на уровне битов как в RAID 2 и 3
- Один диск для контрольной информации
- При операции записи всегда происходит обновление диска четности поэтому только одна операция записи в массив за раз
- Высокая скорость чтения данных больших объемов
- Высокая производительность при большой интенсивности запросов чтения данных
- Достаточно сложная реализация

RAID 5:

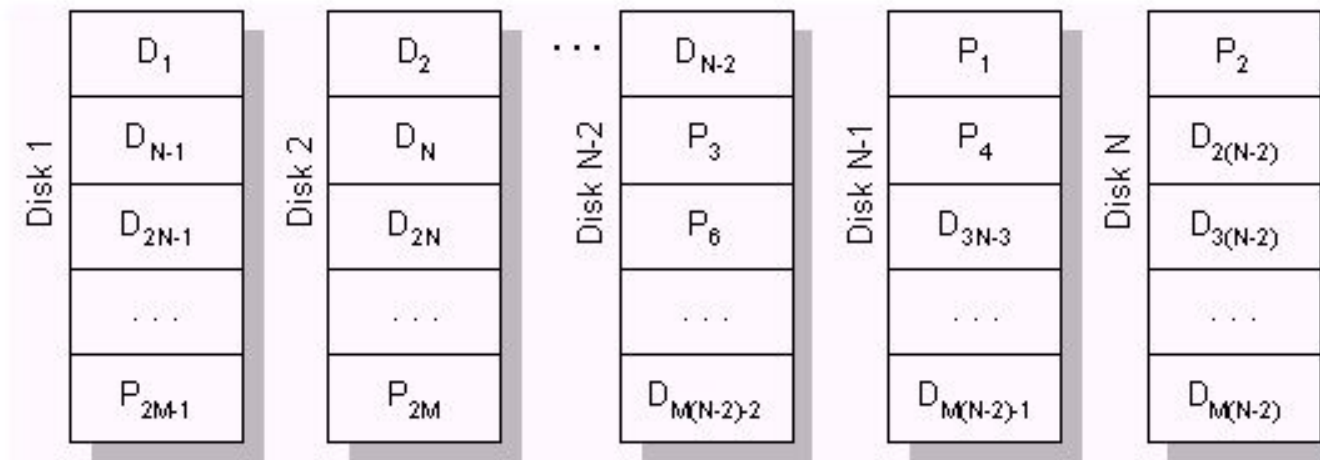
- Контрольная информация чередуется по группе дисков (нет специального диска только для четности как например в RAID 4)
- Параллельные операции записи
- Высокая производительность при большой интенсивности запросов чтения/записи данных
- Скорость чтения данных ниже, чем в RAID 4
- В целом объединяет в себе лучшие черты младших RAID-типов

Магнитные запоминающие устройства

RAID 6:

- Как RAID 5, но с двумя схемами четности
- Высокая отказоустойчивость

Схематическое изображение:



Другие типы RAID:

- Современные RAID контроллеры позволяют комбинировать различные уровни RAID
- RAID 10, RAID 30, RAID 50, RAID 7

Оптические запоминающие устройства

Преимущества в сравнении с магнитными носителями:

- Сменяемость и транспортабельность
- Высокая плотность записи: у DVD 0.4-0.7Гб на кв.см (у магнитных носителей еще выше, но они не сменяемые)
- Меньшая цена в пересчете на мегабайты данных
- Более долгое 'время жизни':
 - Магнитные носители: в районе 3-х лет
 - Оптические носители: от 30 лет (в случае дисков хорошего качества)

Недостатки:

- Больше время доступа
- Более высокая частота появления ошибок; можно обойти использованием кодов с исправлением ошибок

Файловые системы для оптических носителей

Затруднения:

- Стандартные файловые системы создавались для стираемой памяти
- Сложности с адресацией данных большого объема (в частности, более 4Гб)
- Например, файловая система Unix изначально создавалась для работы с жесткими дисками небольшого размера; она слишком медленна при работе с гигабайтами данных

Характеристики файловой системы для оптических носителей:

- Поддержка больших размеров (до десятков терабайт)
- Высокая скорость передачи; низкие издержки на хранение самой файловой системы
- Высокопроизводительная структура каталогов
- Надежность, отказоустойчивость, восстановление при сбоях в каталоге
- Ведение истории и журнала контроля

Файловые системы для оптических носителей

- Файловая система для оптических носителей (ОФС) должна предоставлять избыточную информацию о каталоге; тогда извлечение файлов будет доступно даже в случае серьезного повреждения каталога
- Файлы следует дополнить заголовками, по которым в случае необходимости можно будет перестроить каталог
- Три уровня универсальной ОФС:
 - Верхний уровень: уровень ядра операционной системы
 - Средний уровень: непосредственно сама файловая система (написанная с нуля или модифицированная из файловой системы Unix)
 - Нижний уровень: драйвера устройств (часто связь ОФС с оптическими носителями осуществляется через SCSI-драйвера); драйвера обычно работают только с одной конкретной операционной системой

Файловые системы для оптических носителей

- Файловые системы для магнитных носителей: записи каталога содержатся в специальном файле; последовательный поиск для нахождения нужной записи
- Каталог должен быть выполнен в виде иерархического объекта; вместо последовательного поиска требуются другие способы доступа к записям каталога
- Необходимо хранить более чем одну изолированную копию каталога

Способы поиска по каталогу:

- Хеширование (названий файлов/директорий)
- Модифицированные B-деревья:
 - Узлы большого размера, чтобы уменьшить число уровней
 - Помещение в кэш максимально возможного числа узлов
 - Запись обновленных узлов снизу-вверх
 - Связь в цепочку корневых узлов текущего дерева и старых версий дерева (в случае хранения данных за прошедший период)

Файловые системы для оптических носителей

Методы распределения пространства:

- Таблицы размещения файлов (FAT) не подходят
- 'Ленивое' распределение: высокая скорость при непрерывной передаче данных
- Блоки данных записываются в порядке возрастания
- Размеры блоков – не менее 16-64Кб, для определенных носителей более чем 1Мб
- Поддержка кластеризации; избегать фрагментации
- Высокие фиксированные накладные расходы на каждый доступ к оптическому диску

Чередование данных на нескольких дисках:

- Возможно и для оптических дисков
- Сложность: оптические диски (в дисководов с автоматической сменой дисков) обычно сменные
- Усложнение программного обеспечения

Кэширование и управление томами

Кэширование:

- Энергозависимая (не сохраняющая информацию при выключении питания) ОЗУ
- Постоянный магнитный диск
- Иерархический кэш: ОЗУ плюс жесткий диск
- Алгоритм замещения, например, алгоритм LRU
- Необходимо при 'ленивом' распределении пространства оптического диска
- Обязательно для высокопроизводительных оптических дисковых систем

Управление томами (данных):

- Бд хранит имя тома, расположение, атрибуты и т.д.
- Категории:
 - Оперативные (оперативно-доступные) тома: (подсоединенные диски)
 - Полуоперативные тома: диски в дисководы с автоматической сменой дисков
 - Автономные тома: подсоединяемые вручную

Иерархическое управление запоминающими устройствами

- Генерализация кэширования
- Уровни:
 - 1) ОЗУ;
 - 2) жесткий диск;
 - 3) оперативные/полуоперативные оптические диски;
 - 4) автономные оптические диски
- Задача: увеличение скорости при снижении цены
- Алгоритм: миграция (передвижение по уровням) файлов
 - Доступ к файлу \square помещение файла на высший уровень
 - Файл остается на высшем уровне до тех пор пока есть место или пока пользователи запрашивают данный файл
 - При превышении определенного показателя, некоторые файлы перемещаются на следующий уровень ниже
 - Процесс перемещения охватывает все уровни
 - Предварительная подкачка: перемещение на более высокие уровни может иметь место, если показатель свободного места оказывается ниже определенного уровня (например, вследствие удалений)

Иерархическое управление запоминающими устройствами

Оптимизация для мультимедиа:

- Объединение производительных RAID-массивов с емкостью оперативных/полуоперативных оптических дисководов
- Предварительная загрузка больших видео-клипов на быстрые диски
- Сервисы, предоставляющие видео по требованию, могут базироваться на такой архитектуре

Дальнейшее развитие:

- Постоянно растущие емкость и плотность записи
 - Магнитные носители: теоретический предел плотности записи
 - около 16Гб на кв.см
 - Оптические носители: будет расти до нескольких Гб на кв.см
 - Нанотехнология: продемонстрировано хранение с плотностью около 150Гб на кв.см
- Уменьшение стоимости хранения мегабайта данных