

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Петра Великого**

Институт Компьютерных Наук и Кибербезопасности
Направление: 02.03.01 Математика и компьютерные науки
Высшая Школа Технологий Искусственного Интеллекта

**Отчет по дисциплине: «Архитектура
Суперкомпьютеров»**

«Реализация модели массива
RAID 50, XOR, 10 байт, 10 дисков»

Выполнил: _____
Преподаватель: _____

Салимли А.
Чуватов М.В.

Санкт-Петербург
Весна 2024

Содержание:

Введение	3
1. Определение	4
2. Классификация	5
2.1 RAID 0	5
2.2 RAID 1	5
2.3 RAID 2	5
2.4 RAID 3	6
2.5 RAID 4	6
2.6 RAID 5	6
2.7 RAID 6	7
3. RAID 50	8
4. Заключение	9
5. Список источников	10

Введение

Цель лабораторной работы - реализация модели избыточного массива RAID 50(5+0), с условиями:

- *Дисков в массиве - 10 дисков.*
- *Адреса в массиве - от 0 до 63 (64 адреса).*
- *Размер входных данных - 10 байт (HEX).*
- *Подсчет избыточности - с помощью XOR операции.*

Принцип работы модели:

- *Записывать данные по адресу.*
- *Прочитать данные по адресу.*
- *Восстановление диска при повреждении (удалении).*

1. Определение

RAID (Redundant Array of Independent Disks) - избыточный массив самостоятельных дисков.

Означает, что объединяются два или более дисков в единый элемент с целью повышения отказоустойчивости отдельно взятого элемента массива.

Для дополнительных (избыточных) данных нужен объем. Выделяемый для них объем тоже называется «избыточный».

Само определение избыточный массив - наличие такого объема и чем он больше, тем больше у массива избыточность. Как я сказал ранее, избыточность нужна для отказоустойчивости.

Сам RAID массив - несколько дисков соединенные в единый диск с выделенными местами для хранения информации для исправления ошибок в данных.

Классификация RAID массивов определяются параметрами исполнения RAID-контроллера, по типам поддерживаемых интерфейсов накопителей и по поддерживаемым уровням.

2. Классификация

RAID 0

Этот массив также называется "стримом", потому что при его использовании информация разбивается на фиксированные блоки и записывается поочередно на все имеющиеся диски. Такая технология применяется в библиотеках игр, видеомонтаже или при рендеринге. Минимальное количество дисков для создания RAID 0 - два. Используя данный уровень RAID, скорость чтения и записи информации увеличивается, так как операции выполняются параллельно на всех дисках. Чем больше дисков в массиве, тем выше производительность. Доступен объем всех дисков. Главным недостатком является то, что отказ одного диска в массиве означает полную потерю данных, так как информация хранится разделенной между дисками.

RAID 1

Подходит для хранения важных данных, где безопасность и доступность имеют высший приоритет. Для реализации этого уровня RAID требуется использование двух накопителей. Благодаря дублированию информации обеспечивается высокая надежность системы - при отказе одного диска работа системы будет продолжена с другого. Однако необходимо провести замену поврежденного диска как можно быстрее, при этом данные могут быть восстановлены с зеркального диска. Путем распараллеливания запросов достигается высокая скорость чтения данных. Поскольку диски являются полными копиями друг друга, доступный объем хранилища равен объему одного диска. Логичным дальнейшим шагом является увеличение стоимости хранения на гигабайт данных вдвое. При этом скорость записи информации может оставаться на уровне одиночного диска или даже быть ниже.

RAID 2

Эти массивы применяют чередование дисков и коды коррекции ошибок (код Хэмминга). Благодаря этому диски в них распределяются на две категории: для самих данных и для упомянутых выше кодов.

Благодаря чередованию обеспечивается высокая скорость работы с данными по сравнению с одним диском. А код Хэмминга позволяет обнаруживать и исправлять ошибки при работе с файлами без снижения скорости работы с данными. Кроме того, при отказе одного накопителя массива данные будут восстановлены по сохранённым кодам коррекции ошибок.

RAID 3

Этот тип массива также использует чередование дисков, но без применения кодов Хэмминга. Вместо них используются контрольные суммы, которые используются для восстановления данных, при этом сами данные разбиваются на байты. Такой массив хорошо подходит для работы с большими файлами и потоковым мультимедиа, однако на практике он встречается редко из-за своей низкой надёжности. Для создания такого массива требуется минимум три диска. Скорость чтения данных достаточно высока, а скорость записи — высока только при работе с большими файлами. RAID 3 предлагает хорошее сочетание доступного объёма и стоимости. Информация теряется в случае выхода из строя более одного диска. К недостаткам относится возможное снижение скорости при работе с небольшими объёмами данных, а также то, что не все контроллеры поддерживают работу с таким массивом. Ещё одним минусом является высокая нагрузка на диск, который хранит контрольные суммы, что может привести к его преждевременному

RAID 4

RAID 4 использует блочное чередование с одним диском для четности. Это означает, что данные разбиваются на блоки, и каждый блок записывается на разные диски, а блок четности записывается на отдельный диск. Таким образом, при чтении блока данных из RAID 4 массив может получить все необходимые данные с нескольких дисков одновременно, что увеличивает скорость чтения. Однако запись может замедляться, поскольку все диски должны ждать, пока блок четности будет записан на диск четности перед тем, как следующий блок может быть записан. Обеспечивает хорошую производительность при чтении, но имеет низкую производительность при записи из-за необходимости синхронизации записи блока четности. Он также менее эффективен, чем другие уровни RAID, такие как RAID 5, поскольку использует один дополнительный диск только для хранения четности, что уменьшает общий объем доступного пространства для хранения данных.

RAID 5

Тут уже используется почти такое же чередование как в прошлом массиве, это приводит к увеличению скорости записи, поскольку операции могут выполняться параллельно. В массиве должно быть не менее трёх накопителей. Объём для хранения контрольных сумм

равен объёму одного накопителя. RAID 5 является наиболее популярным и используется в файловых серверах, общих серверах хранения, серверах резервного копирования, работе с потоковыми данными и в других средах, где важна хорошая производительность. Объём накопителей зависит от их количества в массиве – чем больше накопителей, тем больше объём. Скорость чтения высокая, выше, чем у RAID 4, и скорость записи. За счёт распределения сумм по массиву нагрузка на все диски равномерная. Однако, когда один диск выходит из строя, надёжность массива значительно снижается, и он переходит в критическое состояние. Процесс восстановления занимает много времени, что приводит к снижению производительности массива и увеличению нагрузки на накопители, поскольку с них происходит интенсивное продолжительное

RAID 6

Использует два различных блока четности, которые распределены по всем дискам в массиве. Это позволяет восстановить данные даже в случае двух дисков, которые вышли из строя.

Таким образом, RAID 6 обеспечивает высокую степень защиты данных и надёжности.

Однако, использование двух блоков четности требует больше места для хранения данных, что уменьшает общий объем доступного пространства.

Скорость чтения и записи в RAID 6 обычно выше, чем в RAID 5, благодаря параллельной обработке данных на нескольких дисках. Однако, при записи новых данных массиву требуется больше времени для обновления обоих блоков четности, что может снизить общую производительность. Его обычно используют в крупных корпоративных системах хранения данных, где требуется высокая надёжность и защита от сбоев.

3. RAID 50 (5+0)

RAID 50 представляет собой комбинацию уровней RAID 0 и RAID 5. Это означает, что он использует чередование данных и блоков четности, чтобы обеспечить высокую производительность и надежность хранения данных.

RAID 50 состоит из нескольких массивов RAID 5, объединенных в один большой массив с помощью технологии RAID 0. Каждый массив RAID 5 внутри RAID 50 работает независимо, что позволяет сохранять данные даже при отказе одного диска в одном из массивов.

Преимущества RAID 50 включают высокую скорость записи и восстановления данных.

Также, благодаря использованию нескольких массивов RAID 5, RAID 50 может поддерживать большую емкость и обеспечивать высокую доступность данных.

Минимальное количество дисков для создания RAID 50 составляет 6!

На рисунке 1 представлена схема массива.

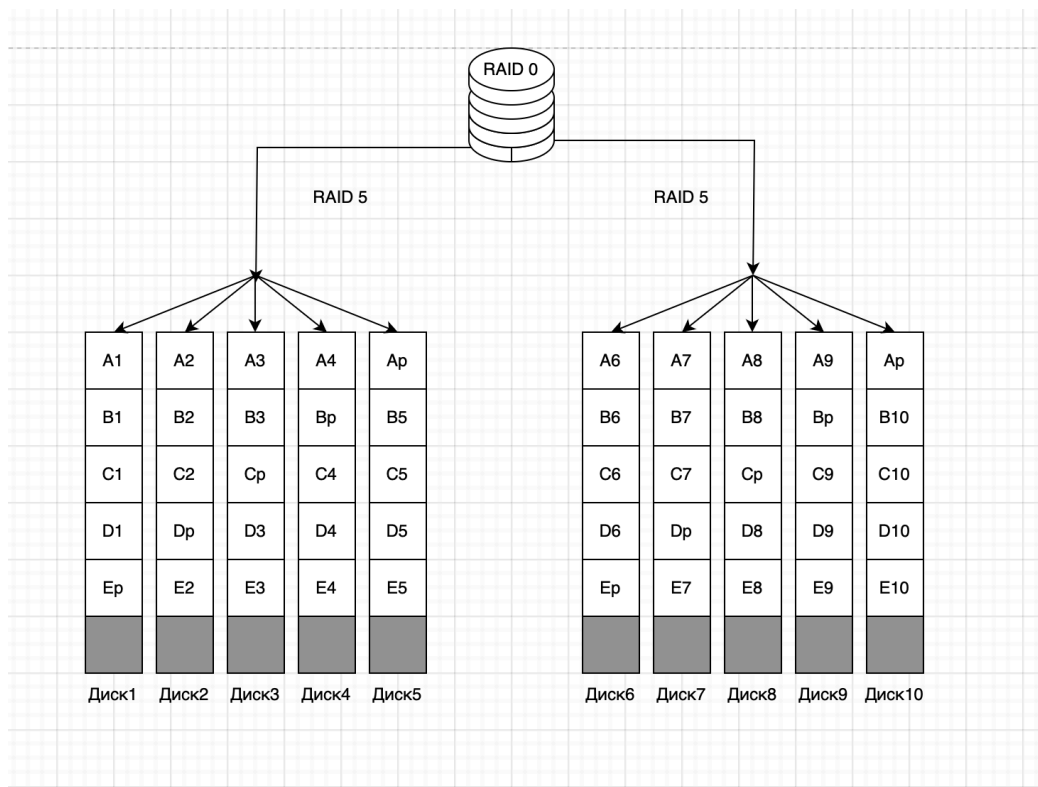


Рис.1 Схема RAID 50 для 10 дисков

Данные разделены на 5 частей: A1, A2, A3, A4, A5. В первую часть массива записывается первая часть данных A1, A2, A3, A4 и рассчитывается избыточность Ap. По такому же принципу записывается вторая часть данных во второй массив: A6-A9 и Ap для этих данных.

4. Заключение

В ходе выполнения лабораторной работы, была реализована модель, которая имитирует работу избыточного массива независимых дисков RAID 50, по условиям, для 10 дисков и для ограничения в 10 символов (HEX), с подсчетом избыточности при помощи XOR операции.

- Были получены знания о принципах работы RAID массивов.
- Была реализована модель работы RAID массива на языке программирования Python 3.9 в среде разработки VS Code.
- Было произведено тестирование нашей модели, посредством выполнения вставки 10-символьного значения HEX, удаления диска, восстановления диска и чтения информации по адресам записи.

Достоинства:

- Можно восстанавливать диски без потери данных
- Возможность чтения по адресу

Недостатки:

- Не самое лучшее количество дисков.

5. Список источников

1. RAID, <https://ru.wikipedia.org/wiki/RAID>, [электронный ресурс], (дата обращения - 21.05.2024г.)
2. RAID-МАССИВЫ, ИХ УРОВНИ, <https://mixtelecom.ru/blog/raid-massiv>, [электронный ресурс], (дата обращения - 21.05.2024г.)
3. Когда используют RAID 50, <https://www.dataarc.ru/articles/when-to-use-raid-50.html>, [электронный ресурс], (дата обращения - 21.05.2024г.)