

Лекция 10. Организация и эксплуатация дисковых подсистем серверов, построенных на основе отказоустойчивых дисковых массивов уровней RAID.

5.2. Отказоустойчивые дисковые массивы серверов

(Краткое изложение содержания учебного пособия)

Для обеспечения отказоустойчивой работы серверов и надежного хранения информации обычно используют отказоустойчивые дисковые массивы с резервированием и/или дублированием дисков, которые называют RAID массивы.

RAID стали расшифровывать как избыточный (резервный) массив независимых дисков «Redundant Array of Independent Disks», потому что для дисковых массивов стали использовать дорогое оборудование, чтобы увеличить не только отказоустойчивость дисковой подсистемы за счет ее избыточности, но и производительность за счет параллельной обработки запросов в «быстрых» дисках.

Современные дисковые массивы RAID дают возможность осуществлять операции записи и чтения фактически одновременно на нескольких дисках.

Основу технологии дисковых массивов RAID составляет избыточность информации, позволяющая в случае сбоя восстановить утраченные данные

Основным компонентом массива RAID является контроллер, который выполняет функцию размещения данных по массиву и позволяет представить весь массив как одно логическое устройство.

Различают следующие основные уровни массивов RAID, которые часто используют в практической работе.

- RAID 0 — дисковый массив повышенной производительности с чередованием, без обеспечения отказоустойчивости;
- RAID 1 — зеркальный дисковый массив;
- RAID 2 — дисковый массив, использующий код Хемминга; (практически не используется)
- RAID 3 и RAID 4 — дисковые массивы с чередованием и выделенным диском чётности;
- RAID 5 — дисковый массив с чередованием и невыделенным диском чётности;
- RAID 6 — дисковый массив с чередованием, с двумя невыделенными дисками, использующий две контрольные суммы, вычисляемые двумя независимыми способами;
- RAID (1+0) — массив RAID 0, построенный из массивов RAID 1;
- RAID (0+1) — массив RAID 1, построенный из массивов RAID 0;
- RAID 51 — массив RAID 1, построенный из массивов RAID 5;
- RAID 30 — массив RAID 0, построенный из массивов RAID 3;
- RAID 50 — массив RAID 0, построенный из массивов RAID 5;
- RAID 60 — массив RAID 0, построенный из массивов RAID 6.

1. Дисковые массивы уровня RAID-0. На самом деле они не являются отказоустойчивыми, так как в них отсутствуют избыточные диски, поэтому отказ любого диска приводит к отказу всего дискового массива.

Однако они имеют наибольшую производительность среди всех дисковых массивов. Контроллер делит поступающую информацию, например, файл на блоки, которые параллельно записывает на отдельные физические диски или считывает с них.

При этом производительность RAID-0, при чтении данных, не всегда увеличивается линейно с числом используемых дисков. Это связано с тем, что у дисков, входящих в дисковый массив, отсутствует синхронизация вращения шпинделей, поэтому вращение каждого диска и поиск информации на нем осуществляется индивидуально, следовательно считывание данных с дисков может идти с некоторым сдвигом.

Минимальное число дисков в этом массиве равно двум.

Дисковые массивы уровня RAID-0 предназначены для систем, требующих большой производительности от дисковых массивов сервера.

2. Дисковые массивы уровня RAID-1 работают по системе дублирования основных дисков, каждый из которых может иметь несколько дублирующих дисков.

Обычно каждый основной диск имеет один дублирующий диск.

Каждый дублирующий диск хранит полную копию основного диска.

При записи данные отдельно записываются на основной диск и на диск, который его дублирует. Если случается сбой или отказ одного из основных дисков, то контроллер заменяет его соответствующим дублирующим диском.

Поэтому отказ всего дискового массива происходит только в том случае, если отказывают основной и дублирующий его диски.

Время записи данных на диски в этих дисковых массивах несколько больше, чем для одного диска.

Минимально допустимое число дисков в дисковом массиве равно двум.

Основной недостаток таких дисковых массивов — высокая цена, так как для их реализации приходится удваивать объем необходимой дисковой памяти.

Дисковые массивы уровня RAID-1 предназначены для систем, требующих высокой надежности дисковых массивов сервера.

3. Дисковые массивы уровня RAID-3 имеют основные диски и только один избыточный, т. е. выделенный дополнительный диск для хранения результатов контроля четности.

При записи данных в этот дисковый массив контроллер выполняет следующие функции:

- разбивает данные (файлы) на блоки (байты);
- формирует из последовательно следующих друг за другом байт группы байт, число которых соответствует числу основных дисков массива;
- параллельно записывает байты каждой группы на разные диски массива;
- для каждой группы записанных байт проводит вычисления на четность и результаты вычислений записывает в виде байта на дополнительный диск.

Дисковые массивы RAID-3 обладают высоким уровнем надежности хранения данных и высокой скоростью их чтения. Скорость записи у RAID-3 меньше, поскольку требуется время для проведения расчетов на четность байт, записываемых на диски.

Для работы массива RAID-3 необходимо минимум три диска: два для хранения данных и один для хранения информации о четности этих данных.

4. Дисковые массивы уровня RAID-4 во многом схожи с массивами RAID-3, также имеют основные диски и только один избыточный, т.е. выделенный дополнительный диск для хранения контроля четности блоков данных, записываемых на основные диски. Контроллер также разбивает данные на блоки, записывает параллельно блоки на основные диски, осуществляет вычисление контроля четности блоков, результаты которого записывает на дополнительный диск.

Однако размер блока в массиве RAID-4 намного больше, чем в RAID-3, он кратен размеру сектора диска и может иметь величину до 128 Кбайт.

При отказе любого диска этого дискового массива его содержимое можно восстановить по информации, которая имеется на остальных исправных дисках.

Отказ дискового массива происходит только в том случае, если происходит отказ любых двух дисков, входящих в его состав.

Дисковые массивы RAID-4 обладают очень хорошей производительностью при работе с данными большого размера, когда параллельно работают все диски. Поэтому, как правило, RAID-4 используют в серверах при решении задач, связанных с обработкой больших массивов данных и небольшой интенсивностью их поступления.

Минимально допустимое число дисков в этом дисковом массиве равно трем.

5. Дисковые массивы уровня RAID-5 имеют основные диски и только один избыточный, но невыделенный дополнительный.

Вычисленный контроль четности блоков, распределяется по всем дискам массива, включая основные и дополнительный.

Здесь нет, в отличие от массивов уровня RAID-3 и RAID-4, четко выделенного диска с контролем четности.

Контроллер разбивает основные данные и данные контроля четности на блоки, которые записывает на разные диски. Это позволяет выполнять несколько операций записи и считывания одновременно.

Дисковые массивы уровня RAID-5 отличаются еще и тем, что их контроллер использует большие размеры блоков данных, записываемых на диск. Поэтому в этих дисковых массивах информация, относящаяся к одному запросу и подлежащая записи или считыванию, как правило, размещается в двух блоках данных: в одном блоке основные данные, а в другом — данные контроля четности.

На обработку каждого запроса задействовано, как правило, два диска.

При выполнении операций записи на диск контроллер может параллельно обрабатывать до $N/2$ запросов, где N — число дисков в дисковом массиве (четное). Если N — нечетное, то число обрабатываемых запросов $(N-1)/2$.

Контроллер также одновременно может выполнять операции записи и считывания на разные диски.

При отказе любого диска этого дискового массива его содержимое можно восстановить по информации, которая имеется на остальных исправных дисках.

Отказ дискового массива происходит только в том случае, если происходит отказ любых двух дисков, входящих в его состав.

Минимально допустимое число дисков в этом дисковом массиве равно трем.

Дисковые массивы уровня RAID-5 в основном ориентированы для работы в многопользовательских системах с небольшими объемами данных, например, для систем обработки транзакций.

5. Дисковые массивы уровня RAID-6 имеют два избыточных, но невыделенных дополнительных диска для ведения двойного контроля четности.

Для каждого блока данных имеем два блока контроля четности, которые вычисляются по разным алгоритмам.

Контроллер разбивает данные на блоки так же, как и в массивах уровня RAID-5. Блоки данных и блоки контроля четности контроллер распределяет по всем дискам массива и записывает их все параллельно на разные диски. Это позволяет контроллеру выполнять операции записи и считывания одновременно.

Следует иметь в виду, что этот дисковый массив имеет очень низкую производительность при записи данных на диск, так как для каждого блока необходимо провести расчет двух блоков контроля четности..

При отказе одного или двух дисков дискового массива его содержимое можно восстановить по информации, которая имеется на остальных исправных дисках.

Отказ дискового массива происходит только при отказе любых трех дисков, входящих в его состав.

Минимально допустимое число дисков в этом дисковом массиве равно четырем.

Дисковые массивы уровня RAID-6 предназначены для систем, которые содержат большие объемы информации и требуют высокой надежности от дисковых массивов сервера.

6. Дисковые массивы уровня RAID-(0+1) относят к классу комбинированных дисковых массивов, поскольку они объединяют и используют скоростные преимущества массива RAID-0 и надежность, присущую массиву RAID-1.

Дисковый массив RAID-(0+1) — это массив, состоящий из двух идентичных массивов RAID-0, которые объединены в массив RAID-1. Поэтому полезный объем массива RAID-(0+1) равен половине суммы объемов всех входящих в массив дисков.

Каждый диск имеет дублирующий диск, на котором находится та же информация, что и на нем.

В случае отказа одного из дисков массив уровня RAID-(0+1) сразу превращается в массив уровня RAID-0.

Отказ дискового массива уровня RAID-(0+1) происходит в том случае, если откажут хотя бы по одному диску, входящих в состав разных массивов уровня RAID-0.

Минимально допустимое число дисков в этом дисковом массиве равно четырем. Число дисков в массиве всегда четное.

Дисковые массивы уровня RAID-(0+1) обычно используют для систем, содержащих небольшие объемы очень важных данных, которые нужно обрабатывать с высокой производительностью.

7. Дисковые массивы уровня RAID-(1+0) объединяют и используют скоростные преимущества массивов уровня RAID-0 и надежность, присущую массивам уровня RAID-1.

Дисковый массив уровня RAID-(1+0) — это массив уровня RAID-0, сегментами которого являются массивы уровня RAID-1.

При его создании сначала диски попарно объединяют в массивы уровня RAID-1, а затем используют для них попеременную запись, как в массивах уровня RAID-0.

Дисковый массив уровня RAID-(1+0) способен работать при отказе до половины дисков массива, практически не теряя своей производительности.

Отказ дискового массива уровня RAID-(1+0) происходит только в случае одновременного отказа двух дисков, образующих зеркальную пару в составе любого массива уровня RAID-1.

Полезная емкость массива уровня RAID-(1+0) составляет 50 % суммарной емкости его дисков.

Минимально допустимое число дисков в этом дисковом массиве равно четырем. Число дисков в массиве всегда четное.

Дисковые массивы уровня RAID-(1+0) обычно используют для систем, содержащих небольшие объемы очень важных данных, которые нужно обрабатывать с высокой производительностью.

8. Дисковые массивы уровня RAID-51 (Иногда их называют RAID-15).

Массив объединяет надежность, присущую массивам уровней RAID-1 и RAID-5. При его создании сначала формируют два массива уровня RAID-5, которые затем объединяют в один массив уровня RAID-1.

Эти массивы имеют не только очень большую отказоустойчивость, но и большую избыточность дисков, а следовательно, и большую стоимость реализации.

Дисковые массивы уровня RAID-51 используют для систем, содержащих небольшие объемы сверхважных данных.

Результаты сравнения наиболее часто используемых дисковых массивов показывают следующее:

По надежности (среднее время безотказной работы дискового массива):
 $H_1 > H_{(1+0)} > H_{(0+1)} \approx H_5 > H_0$ лучший RAID -1

По производительности:
 $P_0 > P_{(1+0)} \approx P_{(0+1)} > P_1 > P_5$ лучший RAID -0

По стоимости :
 $C_{(1+0)} \approx C_{(0+1)} > C_1 > C_5$ лучший RAID -5

Пояснения

- надежность дискового массива уровня RAID-(1+0) немного меньше, чем у RAID-1, но больше, чем у RAID-(0+1) или RAID-5, которые отказывают при отказе любых двух дисков массива;
- скорость работы дискового массива уровня RAID-(1+0) существенно меньше, чем у RAID-0, но намного больше, чем у RAID-1;
- в дисковом массиве уровня RAID-(1+0) информация распределяется строго по зеркалированным дискам, а в массиве уровня RAID-(0+1) имеет место зеркалирование распределенной информации. Поэтому дисковый массив уровня RAID-(1+0) имеет бóльшую надежность, чем дисковый массив уровня RAID-(0+1).

Рекомендации по выбору уровня дискового массива RAID, в зависимости от предъявляемых к нему требований, даны в табл. 5.3

Таблица 5.3

Рекомендации по выбору уровня дискового массива RAID

Первый показатель выбора массива	Второй показатель выбора массива	Рекомендуемый уровень дискового массива RAID*
Производительность	Низкая стоимость Надежность	RAID-0 RAID-(1+0)
Низкая стоимость	Производительность Надежность	RAID-0 RAID-5 или RAID-4
Надежность	Производительность Низкая стоимость	RAID-(1+0) RAID-5 или RAID-4
* Если размер блока меньше размера передаваемой информации, следует выбирать массив уровня RAID-4, если наоборот — массив уровня RAID-5.		

Как правило, для любого дискового массива RAID известны следующие исходные параметры его функционирования:

p — вероятность безотказной работы одного диска;

q — вероятность отказа одного диска, $q = 1 - p$;

d — минимальное число дисков дискового массива RAID, отказ которых приводит к отказу работы дискового массива, для RAID-4 и RAID-5 $d = 2$, для RAID-6 $d = 3$;

n — число дисков в дисковом массиве без RAID;

N — число дисков в составе дискового массива RAID.

Для оценки надежности P (т. е. вероятности безотказной работы дисковых массивов RAID), следует использовать следующие выражения:

для RAID-0 $P = p^n$;

для RAID-1 $P = (1 - (1 - p)^2)^n$;

для RAID-4, RAID-5 и RAID-6 формулу Бернулли, тогда

Для RAID-4, RAID-5
$$P = \sum_{k=0}^1 C_N^k q^k p^{N-k};$$

Для RAID-4, RAID-5
$$P = \sum_{k=0}^2 C_N^k q^k p^{N-k};$$

Для RAID-51 можно считать, что система откажет тогда, когда откажут одновременно два основных диска и два диска, дублирующие эти основные диски, при этом полагаем, что вероятность одновременного выхода из строя пяти и более дисков практически нулевая. $P \approx (1 - (1 - p)^4)^A$, (5.12)

где
$$A = C_{n+1}^2 = \frac{(n+1)!}{(n-1)! \cdot 2!} = \frac{(n+1)n}{2}$$

Сравнительный анализ дисковых массивов уровней RAID по вероятности безотказной работы приведен в табл. 5.4

Таблица 5.4

Значения вероятности безотказной работы дисковых массивов уровней RAID

Значения «n», «p»	RAID-0	RAID-1	RAID-5	RAID-6	RAID-51
	N = n	N = 2n	N = (n + 1)	N = (n + 2)	N = 2(n + 1)
n=2 p=0,9	0,81000	0,98010	0,97200	0,99630	0,99970
n=3 p=0,9	0,72900	0,97030	0,94770	0,99144	0,99940
n=4 p=0,9	0,65610	0,96060	0,91854	0,98415	0,99900
n=5 p=0,9	0,59049	0,95100	0,88574	0,97431	0,99850
n=2 p=0,95	0,90250	0,99501	0,99275	0,99952	0,99998
n=3 p=0,95	0,85737	0,99252	0,98598	0,99884	0,99996
n=4 p=0,95	0,81451	0,99004	0,97741	0,99777	0,99994
n=5 p=0,95	0,77378	0,98756	0,96723	0,99624	0,99991

Считают, что вероятность безотказной работы дискового массива уровня RAID соответствует коэффициенту готовности этого дискового массива.

$$N_i = n + n_{\text{доп}i}$$

Где N_i - количество дисков в составе RAID i -го уровня..

n - количество основных дисков, которые используют для создания RAID.

n_{min} - минимальное количество основных дисков, которое может быть в RAID i -го уровня..

$$n \geq n_{\text{min}i}$$

$n_{\text{доп}i}$ - количество дополнительных дисков, которое следует добавить к основным для создания в RAID i -го уровня..

Таблица 5.5

Состав и особенности RAID систем.

Уровень RAID	n_{min}	$n_{\text{доп}i}$	N_i	Отказ RAID системы
RAID-0	2	0	n	Отказ любого диска RAID
RAID-1	1	n	$2n$	Отказ основного и дублирующего его дисков. RAID
RAID-3	2	1	$n+1$	Отказ любых двух дисков, входящих в RAID
RAID-4	2	1	$n+1$	Отказ любых двух дисков, входящих в RAID
RAID-5	2	1	$n+1$	Отказ любых двух дисков, входящих в RAID
RAID-6	2	2	$n+2$	Отказ любых трех дисков, входящих в RAID
RAID-(0+1)	2	n	$2n$	Отказ хотя бы по одному диску, входящих в состав разных массивов уровня RAID-0.
RAID-(1+0)	2	n	$2n$	Отказ двух дисков, образующих зеркальную пару в составе любого массива уровня RAID-1.
RAID-51	2	$n+2$	$2(n+2)$	Отказ двух пар основного и дублирующего его дисков.