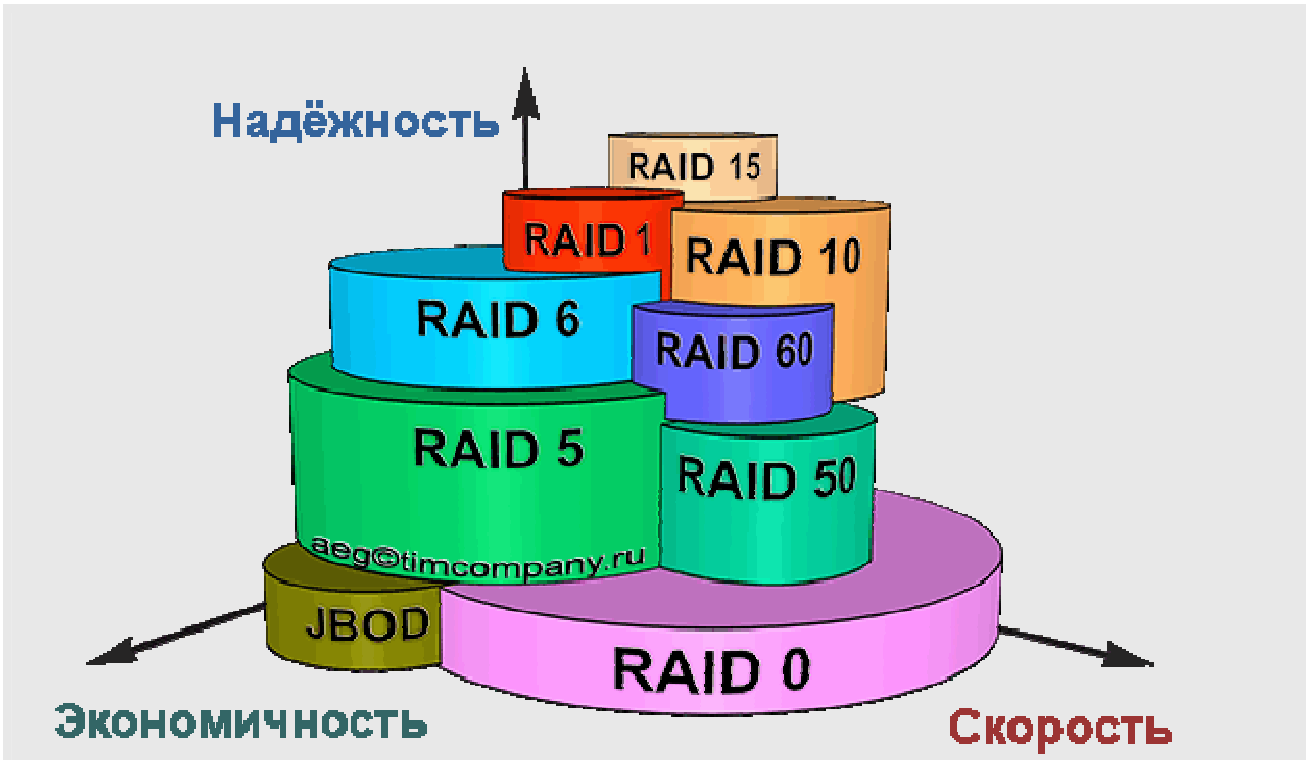
RAID 0, RAID 1, RAID 5, RAID6, RAID 10 или что такое уровни RAID?

Посетители форума задают нам вопрос: «Какой уровень RAID самый надежный?» Все знают, что наиболее распространенным является уровень RAID5, однако он отнюдь не лишен серьезных недостатков, которые неочевидны для неспециалистов.

RAID 0, RAID 1, RAID 5, RAID6, RAID 10 или что такое уровни RAID?

В своей статье я попытаюсь охарактеризовать самые популярные уровни RAID, а затем сформулирую рекомендации по использованию этих уровней. Для иллюстрации статьи я построил диаграмму, на которой поместил эти уровни в трехмерном пространстве надежности, производительности и ценовой эффективности.



Редакция 2010 года:

Экономичность. Следует пояснить, что под *экономичностью* мы подразумеваем в данном случае коэффициент использования пространства жёстких дисков. Для уровней JBOD и RAID0 он равен единице, для "зеркальных" уровней RAID1x равен 1/2, а для всех других вычисляется по формуле (N-X)/N, где X=1 для RAID5, X=2 для RAID5EE и RAID6. При наличии диска (дисков) горячего резерва HotSpare значение X

следует увеличить ещё на количество таких дисков.

Поясню на примере. Для внешней системы хранения данных ProStor® M-6160FA-512, имеющей в составе 16 жёстких дисков, 15 из которых объединены в RAID6, а ещё один объявлен как HotSpare, коэффициент использования будет (16–2–1)/16, то есть 81%. Для массива RAID0 на той же стойке этот коэффициент будет равен 100%, а для RAID1 – 50%. Естественно, ни о какой дешевизне в таком контексте речи быть не может, поэтому не следует анализировать нашу диаграмму на предмет поиска самых бюджетных вариантов.

Раз уж об этом зашла речь, "RAID для бедных" – это, как правило, программные реализации RAID0, RAID1, RAID10, RAID5 на двух, четырёх, максимум - шести дисках. Они чаще всего применяются в домашних станциях и не слишком критических серверных задачах и, конечно, тоже имеют право на существование.

Кстати, термин RAID, предложенный учеными Калифорнийского Университета Беркли в 1987 году, первоначально расшифровывался как Redundant Array of Inexpensive Disks (избыточный массив недорогих дисков). В последующем его изменили на Redundant Array of Independent Disks (избыточный массив независимых дисков) – это более точно отражает его суть и не вводит в заблуждение насчёт стоимости, так как для работы в массиве, естественно, рекомендуется использовать аппаратный контроллер и самые качественные диски, которые дешёвыми не являются.

Скорость. Практически безальтернативным средством долговременного хранения больших объёмов данных до недавнего времени являлись магнитные жесткие диски. Магнитные ленты, различные оптические и магнито-оптические дисковые технологии и флэшки в силу присущих им специфических ограничений составить им конкуренцию не могли. Достойным соперником в будущем могут стать твердотельные диски SSD, но их время ещё не наступило (технологию SSD обсудим чуть позже)).

Итак, "узким" местом любой компьютерной системы являются скоростные характеристики механических частей жёстких дисков: скорость вращения шпинделя и среднее время позиционирования головок. От первой характеристики (при её умножении на плотность записи) зависит так называемая внутренняя скорость чтения и записи данных. От второй – степень затрат времени на перемещение головок, во время которого, естественно, ни запись, ни чтение не производятся.

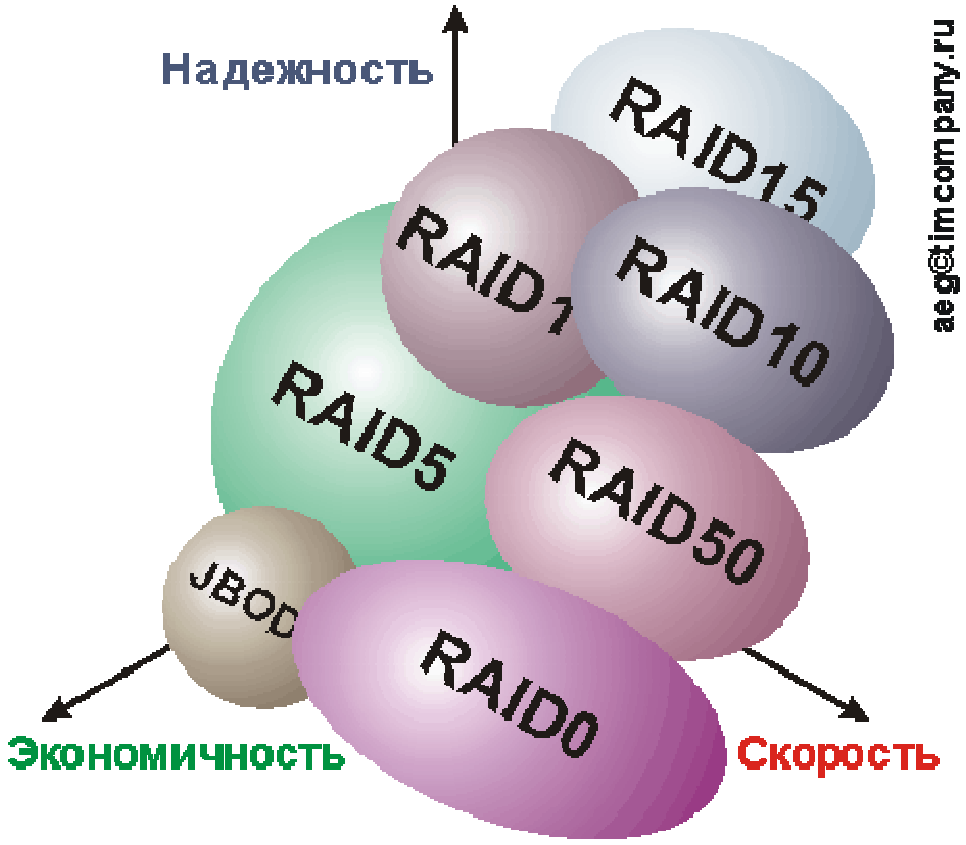
Реальными путями увеличения производительности дисковой подсистемы, очевидно, являются: минимизация перемещения головок накопителей и параллельное чтение данных с нескольких накопителей одновременно. На практике первый путь реализуется увеличением размера кластера операционной системы, оптимизацией размещения данных на диске и

регулярной дефрагментацией. Для реализации второго пути в 1987 было разработано описание набора архитектур массивов дисков, отличающихся отказоустойчивостью и повышенной производительностью, получившее название RAID.

Самыми быстрым согласно нашей диаграмме является уровень RAID0 (RAID00). Теоретически, производительность такого массива должна расти линейно по мере увеличения числа входящих в него дисков, однако на практике, конечно же, всё обстоит не настолько радужно. К тому же следует помнить, что с увеличением числа входящих в RAID0 дисков скорость растёт арифметически, а вероятность отказа – геометрически. В некотором роде компромиссом скорости и надёжности являются все уровни RAIDx0, среди которых упомянем RAID10, RAID1E0, RAID50 и RAID60.

Надёжность. Стопроцентной надежности не бывает в принципе, поэтому в данном случае мы будем рассматривать уровни RAID с точки зрения их устойчивости к отказам отдельных дисков. В этом смысле самым "ненадежным", как мы только что выяснили, является самый быстрый уровень RAID0. Выход из строя любого жёсткого диска в массиве RAID0 вызывает полную потерю его данных.

Самыми надёжными на сегодняшний день можно считать уровни RAID6 и RAID15. В любом случае, не следует забывать про резервное копирование.



JBOD (Just a Bunch of Disks) – это простое объединение (spanning) жестких дисков, которое уровнем RAID формально не является. Томом JBOD может быть массив из одного диска или объединение нескольких дисков. Контроллеру RAID для работы с таким томом не требуется проведение каких-либо вычислений. На нашей диаграмме диск JBOD служит в качестве «ординара» или отправной точки – его значения надежности, производительности и стоимости совпадают с соответствующими показателями единичного жесткого диска.

RAID 0 ("Striping") избыточности не имеет, а информацию распределяет сразу по всем входящим в массив дискам в виде небольших блоков («страйпов»). За счет этого существенно повышается производительность, но страдает надежность. Как и в случае JBOD, за свои деньги мы получаем 100% емкости диска.

Поясню, почему уменьшается надежность хранения данных на любом составном томе – так как при выходе из строя любого из входящих в него винчестеров полностью и безвозвратно пропадает вся информация. В соответствии с теорией вероятностей математически надежность тома RAID0 равна произведению надежностей составляющих его дисков, каждая из которых меньше единицы, поэтому совокупная надежность заведомо ниже надежности любого диска.

Хороший уровень – **RAID 1** ("Mirroring", «зеркало»). Он имеет защиту от выхода из строя половины имеющихся аппаратных средств (в общем случае – одного из двух жестких дисков), обеспечивает приемлемую скорость записи и выигрыш по скорости чтения за счет распараллеливания

запросов. Недостаток заключается в том, что приходится выплачивать стоимость двух жестких дисков, получая полезный объем одного жесткого диска.

Изначально предполагается, что жесткий диск – вещь надежная. Соответственно, вероятность выхода из строя сразу двух дисков равна (по формуле) произведению вероятностей, т.е. ниже на порядки! К сожалению, реальная жизнь – не теория! Два винчестера берутся из одной партии и работают в одинаковых условиях, а при выходе из строя одного из дисков нагрузка на оставшийся увеличивается, поэтому на практике при выходе из строя одного из дисков следует срочно принимать меры – вновь восстанавливать избыточность. Для этого с любым уровнем RAID (кроме нулевого) рекомендуют использовать диски горячего резерва **HotSpare**. Достоинство такого подхода – поддержание постоянной надежности. Недостаток – еще большие издержки (т.е. стоимость 3-х винчестеров для хранения объема одного диска).

Зеркало на многих дисках – это уровень **RAID 10**. При использовании такого уровня зеркальные пары дисков выстраиваются в «цепочку», поэтому объем полученного тома может превосходить емкость одного жесткого диска. Достоинства и недостатки – такие же, как и у уровня RAID1. Как и в других случаях, рекомендуется включать в массив диски горячего резерва HotSpare из расчета один резервный на пять рабочих.

RAID 5, действительно, самый популярный из уровней – в первую очередь благодаря своей экономичности. Жертвуя ради избыточности емкостью всего одного диска из массива, мы получаем защиту от выхода из строя любого из винчестеров тома. На запись информации на том RAID5 тратятся дополнительные ресурсы, так как требуются дополнительные вычисления, зато при чтении (по сравнению с отдельным винчестером) имеется выигрыш, потому что потоки данных с нескольких накопителей массива распараллеливаются.

Недостатки RAID5 проявляются при выходе из строя одного из дисков – весь том переходит в критический режим, все операции записи и чтения сопровождаются дополнительными манипуляциями, резко падает производительность, диски начинают греться. Если срочно не принять меры – можно потерять весь том. Поэтому, (см. выше) с томом RAID5 следует обязательно использовать диск Hot Spare.

Помимо базовых уровней RAID0 - RAID5, описанных в стандарте, существуют комбинированные уровни RAID10, RAID30, RAID50, RAID15, которые различные производители интерпретируют каждый по-своему.

Суть таких комбинаций вкратце заключается в следующем. RAID10 – это сочетание единички и нолика (см. выше). RAID50 – это объединение по “0” томов 5-го уровня. RAID15 – «зеркало» «пятерок». И так далее.

Таким образом, комбинированные уровни наследуют преимущества (и недостатки) своих «родителей». Так, появление «нолика» в уровне **RAID**

50 нисколько не добавляет ему надежности, но зато положительно отражается на производительности. Уровень **RAID 15**, наверное, очень надежный, но он не самый быстрый и, к тому же, крайне неэкономичный (полезная емкость тома составляет меньше половины объема исходного дискового массива).

RAID 6 отличается от RAID 5 тем, что в каждом ряду данных (по-английски *stripe*) имеет не один, а *два* блока контрольных сумм. Контрольные суммы – «многомерные», т.е. независимые друг от друга, поэтому даже отказ двух дисков в массиве позволяет сохранить исходные данные. Вычисление контрольных сумм по методу Рида-Соломона требует более интенсивных по сравнению с RAID5 вычислений, поэтому раньше шестой уровень практически не использовался. Сейчас он поддерживается многими продуктами, так как в них стали устанавливать специализированные микросхемы, выполняющие все необходимые математические операции.

Согласно некоторым исследованиям, восстановление целостности после отказа одного диска на томе RAID5, составленном из дисков SATA большого объема (400 и 500 гигабайт), в 5% случаев заканчивается утратой данных. Другими словами, в одном случае из двадцати во время регенерации массива RAID5 на диск резерва Hot Spare возможен выход из строя второго диска... Отсюда рекомендации лучших RAIDоводов: 1) **всегда** делайте резервные копии; 2) используйте **RAID6**!

Недавно появились новые уровни RAID1E, RAID5E, RAID5EE. Буква “E” в названии означает *Enhanced*.

RAID level-1 Enhanced (RAID level-1E) комбинирует mirroring и data striping. Эта смесь уровней 0 и 1 устроена следующим образом. Данные в ряду распределяются точь-в-точь так, как в RAID 0. То есть ряд данных не имеет никакой избыточности. Следующий ряд блоков данных копирует предыдущий со сдвигом на один блок. Таким образом как и в стандартном режиме RAID 1 каждый блок данных имеет зеркальную копию на одном из дисков, поэтому полезный объем массива равен половине суммарного объема входящих в массив жестких дисков. Для работы RAID 1E требуется объединение трех или более дисков.

Мне очень нравится уровень RAID1E. Для мощной графической рабочей станции или даже для домашнего компьютера – оптимальный выбор! Он обладает всеми достоинствами нулевого и первого уровней – отличная скорость и высокая надежность.

Перейдем теперь к уровню **RAID level-5 Enhanced (RAID level-5E)**. Это то же самое что и RAID5, только со встроенным в массив резервным диском *spare drive*. Это встраивание производится следующим образом: на всех дисках массива оставляется свободным 1/N часть пространства, которая при отказе одного из дисков используется в качестве горячего резерва. За счет этого RAID5E демонстрирует наряду с надежностью лучшую производительность, так как чтение/запись производится параллельно с

бОльшего числа накопителей одновременно и spare drive не простаивает, как в RAID5. Очевидно, что входящий в том резервный диск нельзя делить с другими томами (dedicated vs. shared). Том RAID 5E строится минимум на четырех физических дисках. Полезный объем логического тома вычисляется по формуле N-2.

RAID level-5E Enhanced (RAID level-5EE) подобен уровню RAID level-5E, но он имеет более эффективное распределение spare drive и, как следствие, – более быстрое время восстановления. Как и уровень RAID5E, этот уровень RAID распределяет в рядах блоки данных и контрольных сумм. Но он также распределяет и свободные блоки spare drive, а не просто оставляет под эти цели часть объема диска. Это позволяет уменьшить время, необходимое на реконструкцию целостности тома RAID5EE. Входящий в том резервный диск нельзя делить с другими томами – как и в предыдущем случае. Том RAID 5EE строится минимум на четырех физических дисках. Полезный объем логического тома вычисляется по формуле N-2.

Как ни странно, никаких упоминаний об уровне **RAID 6E** на просторах Интернета я не нашел - пока такой уровень никем из производителей не предлагается и даже не анонсируется. А ведь уровень RAID6E (или RAID6EE?) можно предложить по тому же принципу, что и предыдущий. Диск **HotSpare** *обязательно* должен сопровождать любой том RAID, в том числе и RAID 6. Конечно, мы не потеряем информацию при выходе из строя одного или двух дисков, но начать регенерацию целостности массива крайне важно как можно раньше, чтобы скорее вывести систему из «критического» режима. Поскольку необходимость диска Hot Spare для нас не подлежит сомнению, логичным было бы последовать дальше и «размазать» его по тОму так, как это сделано в RAID 5EE, чтобы получить преимущества от использования бОльшего количества дисков (лучшая скорость на чтении-записи и более быстрое восстановление целостности).

Уровни RAID в «числах».

В таблицу я собрал некоторые важные параметры почти всех уровней RAID, чтобы можно было сопоставить их между собой и четче понять их суть.

Уровень	Избы-точ-ность	Исполь-зование емкости дисков	Произво-дитель-ность чтения	Произво-дитель-ность записи	Встроен-ный диск резерва	Мин. кол-во дисков	Макс. кол-во дисков
RAID 0	нет	100%	**Отл**	**Отл**	нет	1	16
RAID 1	+	50%	Хор +	Хор +	нет	2	2
RAID 10	+	50%	Хор +	Хор +	нет	4	16

RAID 1E	+	50%	Хор +	Хор +	нет	3	16
RAID 5	+	67-94%	**Отл**	Хор	нет	3	16
RAID 5E	+	50-88%	**Отл**	Хор	**+**	4	16
RAID 5EE	+	50-88%	**Отл**	Хор	**+**	4	16
RAID 6	+	50-88%	**Отл**	Хор	нет	4	16
RAID 00	нет	100%	**Отл**	**Отл**	нет	2	**60**
RAID 1E0	+	50%	Хор +	Хор +	нет	6	**60**
RAID 50	+	67-94%	**Отл**	Хор	нет	6	**60**
RAID 15	+	33-48%	**Отл**	Хор	нет	6	**60**

Все «зеркальные» уровни – RAID 1, 1+0, 10, 1E, 1E0.

Давайте еще раз попробуем досконально разобраться, чем же различаются эти уровни?

RAID 1.
Это – классическое «зеркало». Два (и только два!) жестких диска работают как один, являясь полной копией друг друга. Выход из строя любого из этих двух дисков не приводит к потере ваших данных, так как контроллер продолжает работу с оставшимся диском. RAID1 в цифрах: двукратная избыточность, двукратная надежность, двукратная стоимость. Производительность на запись эквивалентна производительности одного жесткого диска. Производительность чтения выше, так как контроллер может распределять операции чтения между двумя дисками.

RAID 10.
Суть этого уровня в том, что диски массива объединяются парами в «зеркала» (RAID 1), а затем все эти зеркальные пары в свою очередь объединяются в общий массив с чередованием (RAID 0). Именно поэтому его иногда обозначают как **RAID 1+0**. Важный момент – в RAID 10 можно объединить только четное количество дисков (минимум – 4, максимум – 16). Достоинства: от "зеркала" наследуется надежность, от «нуля» – производительность как на чтение, так и на запись.

RAID 1E.
Буква "E" в названии означает "Enhanced", т.е. "улучшенный". Принцип этого улучшения следующий: данные блоками "чередуются" ("striped") на все диски массива, а потом еще раз "чередуются" со сдвигом на один диск. В RAID 1E можно объединять от трех до 16 дисков. Надежность соответствует показателям "десятки", а производительность за счет большего "чередования" становится чуть лучше.

RAID 1E0.
Этот уровень реализуется так: мы создаем "нулевой" массив из массивов RAID1E. Следовательно, общее количество дисков должно быть кратно

трем: минимум три и максимум – шестьдесят! Преимущество в скорости при этом мы вряд ли получим, а сложность реализации может неблагоприятно отразиться на надежности. Главное достоинство – возможность объединить в один массив очень большое (до 60) количество дисков.

Сходство всех уровней RAID 1X заключается в их показателях избыточности: ради реализации надежности жертвуется ровно 50% суммарной емкости дисков массива.