# Коррекция скрытых ошибок на жестких дисках – RAID 7.3

***Полина Трофимова,*** *директор по маркетингу RAIDIX*

Принято считать, что наличие системы хранения данных (СХД) гарантирует длительное надежное хранение материала, а при правильном выборе и конфигурировании обеспечивает соблюдение регламента на всех стадиях работы монтажной студии. Действительно, появившись в 1988 году с анонсированием уровней RAID, революционная технология СХД позволила получить недорогие и надежные массивы жестких дисков и инициировала создание отрасли систем хранения данных в ее нынешнем виде. В то время не возникало вопросов о предельных характеристиках частот интерфейсных шин, размерах магнитных ячеек жестких дисков и их влиянии друг на друга, о скоростях вращения шпинделей жестких дисков, потому что в существовавших тогда условиях безошибочную работу с данными гарантировали простые методы коррекции ошибок. С того времени плотность записи на магнитную пластину выросла примерно в 10 тысяч раз, что сделало одной из важных задачу контроля целостности данных и создания различных механизмов коррекции скрытых ошибок, возникающих в процессе передачи и хранения информации. Жесткие диски являются наименее надежным компонентом СХД и требуют дополнительных усилий для обеспечения надежности всей системы. Особое внимание производители уделяют сокращению времени восстановления поврежденных данных, реализации функций выявления и профилактики ошибок данных.

Скрытые ошибки появляются незамеченными на жестких дисках и предоставляются пользователям как верные данные. В случае, когда ошибки выявляются, но не исправляются, результатом становится потеря данных. Даже если скрытая ошибка выявляется, то не всегда можно ее исправить стандартными средствами жесткого диска, и в дело вступают, например, технологии восстановления RAID-массива. Если скрытых ошибок несколько, реконструкция RAID-массива может стать невозможной.

**Массив из 240 жестких дисков в течение пятилетнего периода эксплуатации может нести в себе примерно три скрытые ошибки.**

Как часто возникают скрытые ошибки? Исследования компании NetApp совместно с университетами Висконсина и Торонто показали, что массив из 240 жестких дисков в течение пятилетнего периода эксплуатации может нести в себе примерно три скрытые ошибки, что с точки зрения абсолютных цифр очень мало, но с точки зрения пользователя это может быть катастрофично, так как данные ошибки могут проявиться во время восстановления RAID-массива, когда каждая новая ошибка способна привести к невозможности восстановления данных.

**Каждая новая ошибка способна привести к невозможности восстановления данных.**


Реальность возникновения «порчи» данных подтверждают многие производители компонентов СХД, и этим объясняется использование различных защитных технологий на уровне компонентов и системы в целом.

Все подобные технологии основаны на избыточности данных и призваны значительно уменьшить количество ошибок данных. Как правило, эти защитные технологии рассчитаны на выявление и коррекцию определенного количества ошибочных блоков данных (битов, байтов и т.д.). Если количество блоков ошибочных данных превышает возможности защитной технологии коррекции, то неисправленные ошибки становятся неустранимыми. Существует общепринятый термин BER – Bit Error Rate (Ratio), то есть «уровень невосстанавливаемых ошибок». Применяются еще термины Non-Recoverable Read Error Rate или Unrecoverable (Hard) Data Error Rate. У разных производителей данный параметр может означать 1 sector per 10E16 bits read или 1 per 10E15 bits read.

Под термином «жесткие диски, используемые в СХД», подразумеваются диски корпоративного класса: диски NL SAS (интерфейс SAS, 7200 об/мин) и SAS (интерфейс SAS, 10000…15000 об/мин) и, как исключение, SATA диски корпоративного класса при наличии у производителя СХД технологий, защищающих их от ошибок. Исследование, проведенное специалистами Google (Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz Andr´e BarrosoFailure Trends in a Large Disk Drive Population), указывает на недопустимость использования жестких дисков ниже корпоративного класса в системах хранения.

**Перестроение RAID 6 из дисков по 3 ТВ дает вероятность такой ошибки 72%.**

На сегодняшний день производители СХД применяют в основном диски NL SAS, очевидным преимуществом которых является соотношение цены, емкости и надежности. Согласно данным производителей, вероятность возникновения неустранимой ошибки (BER) у жестких дисков корпоративного класса колеблется в пределах $10^{-15}$ (NL SAS)… $10^{-16}$ (SAS). На практике это означает, что перестроение RAID 6 из 30 дисков по 3 ТВ дает вероятность такой ошибки 72%. Производительность при потоковом чтении современных систем из 60…80 дисков превышает 6 ГБ/с (например, 60-дисковый массив под управлением ПО RAIDIX 3.2 достигает 8 ГБ/с на дисках SATA). Таким образом, при круглосуточном режиме работы неустранимая ошибка чтения может встречаться каждые 6 часов.

**При круглосуточном режиме работы неустранимая ошибка чтения может встречаться каждые 6 часов.**

MTBF (Mean time between failures – среднее время между отказами, наработка на отказ) жестких дисков также влияет на BER. Исследования, проведенные в университете Carnegie Mellon для высоконагруженных систем, показали, что MTBF равно примерно 52,5 ч, то есть выход из строя диска возможен каждые 9 дней.

**Выход из строя диска возможен каждые 9 дней.**

В целях предотвращения потери данных производители выпускают средства для профилактики, выявления и исправления ошибок.

Большинство производителей использует стандарт PI или создают похожие специализированные решения. PI позволяет выявлять ошибки данных, используя дополнительные байты контрольной суммы сектора диска. К сожалению, PI и его аналоги позволяют только выявлять ошибки данных, но не исправлять их. Поэтому производители расширяют функционал выявления ошибок, добавляя функции их исправления. Практический интерес представляют технологии, позволяющие исправлять скрытую ошибку в режиме восстановления RAID-массива с уровнем 6 при неисправности одного диска.

Например, компанией RAIDIX была разработана технология RAID 7.3, позволяющая с использованием трех контрольных сумм (triple parity) в массиве получить возможность исправления скрытых ошибок даже при вышедшем из строя диске. Факт наличия скрытых повреждений выявляется в режиме, когда система самостоятельно проверяет данные. В результате происходит не только обнаружение поврежденных данных, но и их коррекция с применением избыточной информации, хранящейся на остальных дисках массива. Таким образом, вероятность выявления скрытых ошибок в момент чтения данных или (что гораздо критичнее) в момент проведения реконструкции, существенно снижается. При этом для коррекции явных ошибок в RAID 7.3 компании RAIDIX допускается выход из строя до трех дисков с возможностью последующей реконструкции. Важно, что производительность СХД для приоритетных приложений не изменяется, а это является важным преимуществом данной технологии для медиаиндустрии.

Другие компании, производящие СХД, используют для выявления и исправления скрытых ошибок при вышедшем из строя жестком диске так называемую «горизонтальную» избыточность – контрольные суммы по блокам данных на каждом диске. Недостатком такого и других подходов по сравнению с решением компании RAIDIX 7.3 является повышение объемов служебной информации и увеличение времени исправления ошибки.

Профилактика (предотвращение) неисправимых и скрытых ошибок выполняется путем заблаговременного сканирования (общепринятое наименование процесса – scrubbing) жесткого диска на предмет нахождения скрытых ошибок до момента, когда эти ошибки окажутся неустранимыми. Такое сканирование в реальном времени может выполняться во время почти любого рабочего процесса жесткого диска: чтения, записи или просто простоя. Компания RAIDIX разработала функцию Silent Data Corruption Protection, которая выполняет постоянное сканирование дисков на предмет выявления скрытых ошибок. Далее с применением собственных технологий производится восстановление данных.

Важным отличием методов предотвращения, выявления и коррекции ошибок компании RAIDIX является отсутствие потерь производительности при выполнении этих операций. Благодаря высоким скоростям расчета контрольных сумм для RAID 6 и RAID 7.3 процедура выявления ошибок не влияет на чтение и запись данных.