

## **Утилита Cloper**

# **Сжатие и оптимизация откатываемого диска виртуальной машины**

**Copyright (C) Антон Копьев,  
GNU General Public License**

## **Введение.**

Утилита Cloper является консольной программой для оптимизации виртуальных дисков отката с целью уменьшения размера их файла на физическом диске. На 32-х и 64-х разрядных ОС возможно использование x86 версии утилиты, в случае 64-х разрядной гостевой ОС возможно использование x64 версии утилиты. Утилита тестировалась на VMWare Workstation v.8.0, но её схема работы должна быть применима ко всем вариантам плееров виртуальных машин, если они не инициализируют весь выделенный объём диска при его создании, а только осуществляют хранение и запись фактических данных. Если же при первичной инициализации виртуального диска плеер машины выделяет под него столько места, сколько указано для объема диска, то использование данной утилиты становится бессмысленным.

Использование утилиты осуществляется на виртуальной машине с ОС Windows, к которой подключены исходный оптимизируемый и целевой эталонный диск в откатываемом режиме. Исходный и целевой диски для создания оптимизированного снимка должны быть копиями одного и того же виртуального диска, при этом целевой диск содержит только первоначальные эталонные данные без добавленной информации в исходном диске.

В целом, процедура оптимизации заключается в сравнении данных двух копий диска с низкоуровневой обработкой данных исходного диска для того, чтобы физические данные его файлов по возможности располагались напротив таких же физических данных на целевом эталонном диске. После выполнения процедур сопоставления осуществляется клонирование данных исходного диска на эталонный. Потому как при клонировании производится только запись данных, которые отличаются от содержимого целевого эталонного диска, то в результате его размер файла на физическом диске значительно сокращается.

Если предполагается дальнейшее использование исходного виртуального диска, то рекомендуется выполнять оптимизацию, сохранив его копию. Характер действий с его данными является обычной фрагментацией и вероятность их повреждения не выше вероятности повреждения файла при копировании в другую папку, но из-за внутреннего перемещения данных происходит инициализация неиспользованных ранее секторов виртуального диска и в результате размер его физического файла вырастает до фактического объёма диска.

## **Процедуры оптимизации.**

Процедуру оптимизации можно применять к виртуальным дискам с форматами размещения разделов MBR и GPT. Операция сопоставления и прочие действия оптимизации применяются только к разделам с файловой системой NTFS, разделы с другими файловыми системами копируются как есть. В утилите предусмотрены следующие процедуры оптимизации NTFS томов:

1. Предварительное удаление указанных для этого файлов источника, исключение из процедур одинаковых файлов с одинаковым расположением кластеров в источнике и в эталоне;
2. NTFS сжатие новых и измененных данных с их консолидацией (то есть, если указано, то также перемещает эти исходные данные перед эталонными данными без совпадений);
3. Поиск совпадений кластеров файлов источника среди не сопоставленных кластеров файлов эталона и перемещение этих данных файлов источника напротив совпавших кластеров эталона;
4. Запись ненулевых секторов эталона в свободные от файлов области тома источника;
5. Потому как при клонировании производится перенос только ненулевых секторов, то остальные "ненулевые" свободные секторы тома источника обнуляются.

## ***Прerequisites.***

Чтобы иметь возможность выполнять задачу оптимизации, гостевая ОС должна обеспечивать доступ к целевым и исходным дискам с использованием функциональных возможностей Windows API, активный пользователь ОС операционной системы должен иметь права администратора. Утилита может быть запущена с использованием версий ОС Windows, начиная с XP SP2 x86 / x64. Рекомендуется использовать более старые версии ОС для минимизации фоновых изменений данных системой во время оптимизации.

Независимо от параметризации задачи, утилита активно работает с данными исходного диска во время всех подзадач. Хост-система должна быть способной поддерживать эту активность в течение всей её работы. Для обеспечения оптимальной производительности рекомендуется хранить образ исходного диска на отдельном жестком приводе.

В зависимости от производительности гостевой ВМ, количества исходных и целевых данных, спецификации задачи, вся процедура может занять от 10-20 минут до нескольких часов. Результирующий оптимизированный образ диска может быть получен только после завершения всех заданных процедур. Чтобы избежать фоновой инициализации оптимизированных данных, рекомендуется выключать гостевую ВМ или отсоединять целевой диск сразу после завершения задачи.

В зависимости от параметризации задачи системные требования к гостевой виртуальной машине могут быть минимальными или могут требовать значительного объема ОЗУ. Минимальный размер ОЗУ для запуска гостевой ОС является необходимым условием для всех задач, кроме подзадачи, которая ищет совпадения исходных данных внутри целевых данных. Последняя подзадача может потребовать до 2 ГБ оперативной памяти только для процесса утилиты, она также значительно потребляет время ЦПУ для поиска соответствий данных.

## ***Parameters of the utility at startup.***

Утилита Cloper является консольной программой с параметрами командной строки и с конфигурационным файлом. Ее вызов без параметров в командной строке выводит их список с описанием. Предусмотрены следующие параметры командной строки:

1. `"/help"` или `"/h"` - вывод справки использования утилиты;
2. `"/getconfig:short"` или `"/gc:short"` - вызов для создания конфигурационного файла с параметрами по умолчанию. Суффикс `":short"` является опциональным, если его не указывать, то будет создан конфигурационный файл с комментариями к каждому параметру и его значению;
3. `"/getserials:<drive letter>"` или `"/gs:<drive letter>"` - получает серийный идентификатор диска, который состоит из подстрок идентификаторов всех его томов разделенных символом "-". Для того, чтобы можно было получить идентификатор диска, он должен иметь по крайней мере один том с назначенной ему буквой. Потому как процедура оптимизации выполняется с использованием двух копий одного и того же диска, то идентификаторы их томов совпадают;
4. `"/serials:<serials>"` или `"/s:<serials>"` - запуск задачи для указанной пары дисков;
5. `"/serials:<serials> /showdiskdirs|sdd:<folders filter>"` или `"/[s]:<serials> /[sdd]:<folders filter>"` - вывод списка согласно указанному фильтру. Этот специальный вызов служит для дополнительного контроля адекватности указанного фильтра отбора папок ожидаемому результату. Значение фильтра отбора объектов для их предварительного удаления указывается в конфигурационном файле в его секции `"[VolumeCleanup]"`. Смотрите справку `"/h"` с номерами разделов 7-9 для дополнительных деталей;
6. `"/serials:<serials> /showdiskfiles|sdd:<files filter>"` или `"/[s]:<serials> /[sdf]:<files filter>"` - вызов аналогичен предыдущему, но для проверки отбора файлов;

7. "[signdisk]:<new signature>" или "[sd]:<new signature>" - интерактивное изменение сигнатуры диска. Если новое значение не указано, то выводит список подключенных дисков с их сигнатурами.

При создании конфигурационного файла его значения устанавливаются в оптимальное значение, в случае создания файла с комментариями он имеет построчную справку. При первом запуске утилиты для задач оптимизации на виртуальной машине рекомендуется создавать конфигурационный файл заново. Для запуска утилиты из скриптов автоматизации параметр "RunMode" должен быть изменен на "SILENT", а параметр "PauseAtEnd" на "NOT".

### **Тестирование и его результаты.**

Изменение параметров влияет на эффективность сжатия файла диска отката. Для демонстрации были произведены прогоны утилиты с одними и теми же данными, но при изменении параметров. Образец данных для сравнения содержал откатываемый снимок диска, размер его файла составлял 52,4 ГБ. Виртуальный диск имел внутренний размер 75 ГБ и два основных раздела. Первый небольшой раздел BCD Windows был исключен из процедур оптимизации, второй раздел содержал установленную ОС Window 10 x64. На целевом эталонном диске не было других данных помимо установленной ОС и нескольких программ, на исходном диске имелось около 40 ГБ пользовательских данных на втором разделе. Результаты тестирования в таблице ниже были получены с использованием виртуальной машины под управлением ОС Windows 7 x64:

Usage of subtask for optimization				Compress Data (NTFS)	Time spent on task (hours)	The size of the snapshot file after tasks (GB)	Notes
Patch Free Space	Move New Data	Match Blocks	Move To Blocks				
#5	#4	#1-2	#3				(1)
NOT	NOT	NOT	NOT	n/a	0.87	35.2	
YES	NOT	NOT	NOT	n/a	1.65	33.8	
YES	NOT	YES	NOT	NOT	2.72	33.0	(2)
YES	NOT	YES	YES	NOT	2.33	33.0	
YES	YES	YES	YES	NOT	2.57	32.7	
YES	NOT	YES	YES	YES	2.97	31.4	
YES	YES	NOT	NOT	YES	2.80	22.1	
YES	YES	YES	NOT	YES	3.55	21.3	(2)
YES	YES	YES	YES	YES	3.47	21.1	

(1) Обработка "Move To Blocks" соответствует перемещению новых данных напротив не сопоставленных данных эталона (поля "MoveNewData" и "ConsolidateNewData" конфигурационного файла).

(2) Увеличение времени выполнения задачи в этих случаях вызвано значительной фрагментацией свободного пространства источника после сопоставления исходных и целевых блоков данных.

Следует отметить, что таблица содержит только результаты с покластерным сравнением блоков при выполнении подзадач № 1-2. Посегментное сравнение занимает больше времени и оперативной памяти во время этих подзадач. Сравнение сегментов имеет смысл использовать только в том случае, если хранение данных гостевой системы (ВМ) в системе хоста (ПК) основано на блоках данных, имеющих размер сектора. По крайней мере в случае VMware хранение гостевых данных скорее всего осуществляется в блоках размером по 4 Кбайт и совпадает со стандартным размером кластера. Случаи разных размеров кластеров файловых систем гостевого тома диска и тома диска хоста не

проверялись, то есть в данном случае возможны те или иные побочные эффекты и общий размер найденных сопоставлений может не совпадать с уменьшением размера файла образа виртуального диска.

Из-за фоновой активности системы разница в окончательном размере файла становится значимой только начиная с 100-200 МБ. Виртуальная машина, работающая под управлением ОС Windows XP SP2, дает аналогичные результаты с небольшим улучшением из-за более низкой фоновой активности системных драйверов. Таким образом, для лучшего результата оптимизации рекомендуется использовать виртуальную машину с более старыми версиями ОС и с минимальным комплектом стороннего программного обеспечения.

Также, для получения наилучшего эффекта сжатия рекомендуется в диске эталоне держать набор больших файлов (например, каких-нибудь архивов), которые можно удалить либо при начале работы пользователя с диском источника, либо при запуске процедур оптимизации утилитой. Если размер этих файлов примерно соответствует объему позже добавленных пользователем данных, то все новые данные источника скорее всего будут полностью сопоставлены этим ненужным данным эталона и размер оптимизированного физического файла может оказаться до 1000 раз меньше размера исходного физического файла виртуального диска. Если плеер виртуальной машины позволяет, то для этих целей может быть полезно использовать вторичный снимок диска с какой-нибудь старой версией исходных данных, которую можно использовать в качестве эталона для сопоставлений и оптимизации. Для последующего использования в данном случае лучше сохранить какую-либо старую версию оптимизируемых данных в больших файлах архивов, при этом архивирование должно производиться без сжатия, без применения NTFS сжатия к этим архивам и их физическое размещение на диске должно быть консолидировано.

### ***Заметки разработчика.***

Первоначальной целью утилиты является её использование в составе системы автоматизации предприятия. Согласно проверкам, возможности современного настольного ПК позволяют иметь серверную систему со средней производительностью 20-30 оптимизированных файлов в день. В зависимости от спецификации автоматизируемой системы, реализация и внедрение занимает 1-3 месяца.

Согласно GNU GPL исходные коды утилиты могут использоваться в любых других программных продуктах под произвольной лицензией. Опубликованная здесь версия программы имеет рабочее состояние и может выполнять описанные действия с данными конечных пользователей в соответствии с приведенными здесь спецификациями. Так как при работе утилиты могут иметь место аппаратные и программные сбои, то рекомендуется выполнять резервные копии перед запуском программы.

Утилита версии 1.12 является первым релизом этой программы для использования, предыдущие версии не публиковались и использовались только для внутренних целей. Текущая версия программы имеет следующие ограничения и возможные направления для дальнейшего развития:

1. Утилита может осуществлять оптимизацию томов только с файловой системой NTFS;
2. При поиске соответствий среди целевых данных утилита на данный момент выбирает варианты с наибольшим числом совпадений для каждого файла. Согласно проверкам это дает результат, близкий к оптимальному распределению соотношения совпадений, но в редких случаях это может давать решение, соответствующее неправильному экстремуму со значительно сниженным общим результатом;
3. Текущие функциональные возможности Windows API не позволяют делать преднамеренную фрагментацию файлов, это приводит к невозможности установки и переноса соответствий для фрагментированных целевых файлов и, в некоторых случаях, для целевых файлов, которые имеют меньший размер на целевом диске;

4. Геометрии исходного и целевого дисков должны совпадать, утилита не может выполнять сопоставление данных, если исходные и целевые тома имеют различия;
5. Утилита не может выполнять дефрагментацию системных файлов NTFS и не может изменять геометрию целевого диска после завершения задачи, это также не позволяет очищать исходные блоки от системных данных;
6. Задача оптимизации утилиты получает результат путем записи данных на эталонный целевой диск.

Помимо того факта, что изменение любого из этих пререквизитов может потребоваться для какой-либо конкретной задачи, существует также множество других возможных смежных областей применения. В результате я не планирую делать какие-либо изменения в текущем варианте программы в ближайшее время. По крайней мере, до получения точной информации от конечных пользователей, касающейся необходимых для использования изменений и условий для поддержки проекта.

### **Файлы проекта утилиты.**

Дополнительные файлы к утилите Cloper добавлены в целях её лучшего соответствия критериям программного обеспечения с открытым кодом, для её документации и, в случае необходимости, для облегчения работы с исходным кодом. Список файлов с их относительным расположением:

1. ".\docs\Утилита Cloper - сжатие и оптимизация диска отката виртуальной машины.pdf" - описание утилиты на русском языке;
2. ".\docs\Cloper utility for compressing and optimizing virtual machine rollback disk.pdf" - описание утилиты на английском языке;
3. ".\project\..." - исходный проект утилиты на Паскале, отладка и компилирование опубликованной версии программы осуществлялось с помощью Embarcadero® RAD Studio XE5;
4. ".\cloper\_x86.exe" - 32-х разрядная версия утилиты;
5. ".\cloper\_x64.exe" - 64-х разрядная версия утилиты.

### **Состояние проекта, его поддержка и развитие.**

Текущий набор функциональности утилиты и исходный код утилиты работоспособен и может использоваться на ОС Windows от версии XP до версии 11. Этот функциональный набор опубликован здесь для общего использования как есть. Для решения других схожих проблем обработки исходных данных виртуальных машин набор возможностей утилиты может быть расширен или доработан. Для предложений по финансовой поддержке и по развитию проекта я доступен по следующим контактам:

E-Mail: kopyurff@yahoo.com, kopyurff@rambler.ru  
Мобильный: 8-921-912-44-10