ISSN 2306-1561
Automation and Control in Technical Systems (ACTS)
2014, No 4, pp. 158-167.
DOI: 10.12731/2306-1561-2014-4-16

Research of File System Capacity for Linux Kernel

Salniy Alexander Gennadievich

Russian Federation, Undergraduate Student, Department of «Automated Control Systems».

State Technical University – MADI, 125319, Russian Federation, Moscow, Leningradsky prospekt, 64. Tel.: +7 (499) 151-64-12. http://www.madi.ru

alexandr.salny@yandex.ru

Ostroukh Andrey Vladimirovich

Russian Federation, full member RAE, Doctor of Technical Sciences, Professor, Department of «Automated Control Systems».

State Technical University – MADI, 125319, Russian Federation, Moscow, Leningradsky prospekt, 64. Tel.: +7 (499) 151-64-12. http://www.madi.ru.

ostroukh@mail.ru

Abstract. The article describes the most common file system Linux kernel. The research was carried out on a personal computer, the characteristics of which are written in the article. On a personal computer for measuring the file capacity, has been installed the necessary software. Based on the results, conclusions and proposed recommendations for use of file systems.

Keywords: file system, operating system (OS), file, capacity, Linux, experimental data, personal computer (PC), software.

ISSN 2306-1561
Автоматизация и управление в технических системах (АУТС)
2014. – №4. – С. 158-167.
DOI: 10.12731/2306-1561-2014-4-16

УДК 004.9

Исследование производительности файловых систем ядра Linux

Сальный Александр Геннадьевич

Российская Федерация, магистрант кафедры «Автоматизированные системы управления».

ФГБОУ ВПО «Московский автомобильно-дорожный государственный технический университет (МАДИ)», 125319, Российская Федерация, г. Москва, Ленинградский проспект, д.64, Тел.: +7 (499) 151-64-12, http://www.madi.ru

alexandr.salny@yandex.ru

Остроух Андрей Владимирович

Российская Федерация, академик РАЕ, доктор технических наук, профессор кафедры «Автоматизированные системы управления».

ФГБОУ ВПО «Московский автомобильно-дорожный государственный технический университет (МАДИ)», 125319, Российская Федерация, г. Москва, Ленинградский проспект, д.64, Тел.: +7 (499) 151-64-12, http://www.madi.ru.

ostroukh@mail.ru

Аннотация. В статье рассмотрены наиболее распространённые файловые системы ядра Linux. Исследование производилось на типовом персональном компьютере, характеристики которого приведены в статье. На персональном компьютере для проведения замеров скорости работы с файлами было установлено необходимое программное обеспечение. На основе полученных результатов, сделаны выводы и предложены рекомендации к применению файловых систем.

Ключевые слова: файловая система, операционная система (ОС), файл, производительность, Linux, экспериментальные данные, персональный компьютер (ПК), программное обеспечение.

1. Введение

Фа́йловая систе́ма – порядок, определяющий способ организации, хранения и именования данных на носителях информации в компьютерах, а также в другом электронном оборудовании: цифровых фотоаппаратах, мобильных телефонах и т. п. [1]

Файловая система определяет формат содержимого и способ физического хранения информации, которую принято группировать в виде файлов. Конкретная файловая система определяет размер имен файлов и (каталогов), максимальный возможный размер файла и раздела, набор атрибутов файла. Некоторые файловые системы предоставляют сервисные возможности, например, разграничение доступа или шифрование файлов.

2. Обзор файловых систем Linux

В гастоящее время наиболее распространены следующие файловые системы: ext2, ext3, ext4, ReiserFS, btrfs, jfs [1 – 20].

Second Extended File System (дословно: «вторая расширенная файловая система»), сокращённо **ext2** – файловая система ядра Linux. Была разработана Реми Кардом взамен существующей тогда ext. По скорости и производительности работы она может служить эталоном в тестах производительности файловых систем.

Главный недостаток ext2 (и одна из причин демонстрации столь высокой производительности) заключается в том, что она не является журналируемой файловой системой. Он был устранён в файловой системе ext3 – следующей версии Extended File System, полностью совместимой с ext2. Но для ssd это скорее плюс, продлевает жизнь накопителя. Это основная причина, почему EXT2 до сих пор поддерживается в Anaconda и Ubiquity.

Файловая система ext2 по-прежнему используется на флеш-картах и твердотельных накопителях (SSD), так как отсутствие журналирования является преимуществом при работе с накопителями, имеющими ограничение на количество циклов записи.

Third extended file system (третья версия расширенной файловой системы), сокращённо **ext3** или **ext3fs** – журналируемая файловая система, используемая в операционных системах на ядре Linux, является файловой системой по умолчанию во многих дистрибутивах. Основана на ФС ext2, начало разработки которой положил Стивен Твиди.

Основное отличие от ext2 состоит в том, что ext3 журналируема, то есть в ней предусмотрена запись некоторых данных, позволяющих восстановить файловую систему при сбоях в работе компьютера.

Стандартом предусмотрено три режима журналирования:

- writeback: в журнал записываются только метаданные файловой системы, то есть информация о её изменении. Не может гарантировать целостности данных, но уже заметно сокращает время проверки по сравнению с ext2;
- ordered: то же, что и writeback, но запись данных в файл производится гарантированно до записи информации об изменении этого файла. Немного снижает производительность, также не может гарантировать целостности данных (хотя и увеличивает вероятность их сохранности при дописывании в конец существующего файла);

– journal: полное журналирование как метаданных ФС, так и пользовательских данных. Самый медленный, но и самый безопасный режим; может гарантировать целостность данных при хранении журнала на отдельном разделе (а лучше — на отдельном жёстком диске).

Файловая система ext3 может поддерживать файлы размером до 1 ТБ. С Linux-ядром 2.4 объём файловой системы ограничен максимальным размером блочного устройства, что составляет 2 терабайта. В Linux 2.6 (для 32-разрядных процессоров) максимальный размер блочных устройств составляет 16 ТБ, однако ext3 поддерживает только до 4 ТБ

Максимальное число блоков для ext3 равняется 232. Размер блока может быть различным, что влияет на максимальное число файлов и максимальный размер файла в файловой системе

Таблица 1 – Ограничения размеров в файловой системе ext3

Размер блока	Макс. размер файла	Макс. размер файловой системы
1 KB	16 GB	до 2 TB
2 KB	256 GB	до 8 TB
4 KB	2 TB	до 16 TB
8 KB]	2 TB	до 32 TB

Ext4 – это результат эволюции Ext3, наиболее популярной файловой системы в Linux. Во многих аспектах Ext4 представляет собой больший шаг вперёд по сравнению с Ext3, чем Ext3 была по отношению к Ext2 [2]. Наиболее значительным усовершенствованием Ext3 по сравнению с Ext2 было журналирование, в то время как Ext4 предполагает изменения в важных структурах данных, таких как, например, предназначенных для хранения данных файлов.

Одним из улучшений является увеличения максимального размера одного файла для данной фс. На сегодняшний день максимальный размер файловой системы Ext3 равен 16 терабайтам, а размер файла ограничен 2 терабайтами. В Ext4 добавлена 48-битная адресация блоков, что означает, что максимальный размер этой файловой системы равен одному экзабайту, и файлы могут быть размером до 16 терабайт. 1 EB (экзабайт) = 1,048,576 TB (терабайт), 1 EB = 1024 PB (петабайт), 1 PB = 1024 TB, 1 TB = 1024 GB. 48-битная адресация блоков применяется из-за ряда ограничений, которые необходимо было бы снять, чтобы сделать Ext4 полностью 64-битной, и такой задачи перед Ext4 не ставилось. Структуры данных в Ext4 проектировались с учётом требуемых изменений, поэтому однажды в будущем поддержка 64 бит в Ext4 появится. Пока же придётся довольствоваться одним экзабайтом.

Так же одной из важных особенностей данной файловой системы является онлайн-дефрагментация, которое позволяет снизить фрагментацию файлов. Например: вы создаёте три файла в одном каталоге и они расположены на диске друг за другом. Потом,

однажды вы решаете обновить второй файл, и при этом файл становится несколько больше – так, что места для него становится недостаточно. При этом нет никаких других решений, кроме как отделить не вмещающийся фрагмент файла и положить его на другое место диска или выделить файлу последовательную область диска большего размера в другом месте, вдалеке от первых двух файлов, что приведёт к перемещениям головки диска, если приложению потребуется считать все файлы в каталоге (скажем, менеджер файлов будет создавать эскизы для файлов изображений). Помимо этого, файловая система может заботиться только об определённых типах фрагментации и она не может знать, например, что она должна хранить все файлы, требуемые при загрузке, рядом друг с другом, поскольку она просто не знает, какие из них требуются при загрузке.

XFS – высокопроизводительная журналируемая файловая система, созданная компанией Silicon Graphics для собственной операционной системы IRIX. 1 мая 2001 года Silicon Graphics выпустила XFS подGNU General Public License. XFS отличается от других файловых систем тем, что она изначально была рассчитана для использования на дисках более 2 терабайт. Поддержка XFS была включена в ядро Linux версий 2.4 (начиная с 2.4.25, когда Марчело Тозатти посчитал её достаточно стабильной) и 2.6, и, таким образом, она стала довольно универсальной для Linux-систем. Инсталляторы дистрибутивов openSUSE, Gentoo, Mandriva, Slackware, Ubuntu, Fedora и Debian предлагают XFS как вариант файловой системы для установки. FreeBSD стала поддерживать XFS в режиме чтения в декабре 2005 года.

Однако данная файловая система обладает весьма критичными для работы сервера недостатками.

- Невозможно уменьшить размер существующей файловой системы.
- Восстановление удалённых файлов в XFS — очень сложный процесс, поэтому на данный момент (2014 год) для этого существует всего лишь несколько программных продуктов, например «Raise Data Recovery for XFS» для ОС Windows.
- Возможность потери данных во время записи при сбое питания, так как большое количество буферов данных хранится в памяти при том что метаданные записываются в журнал (на диск) оперативно. Это характерно и для других файловых систем с журналированием метаданных.

Btrfs (B-tree FS, «Better FS» или «Butter FS») – файловая система для Linux, основанная на структурах Б-деревьев и работающая по принципу «копирование при записи» (copy-on-write). Опубликована компанией Oracle Corporation в 2007 году под лицензией GNU General Public License (GPL). Одной из первоначальных целей разработки данной файловой системы было обеспечение достойной конкуренции популярной ZFS. Btrfs будет избавлена от многих недостатков, присущих другим современным файловым системам для Linux.

Btrfs считается стабильной, однако по состоянию на 2010 год не создано инструмента для проверки файловой системы и исправления ошибок. Версия Btrfs v0.19 выпущена в июне 2009 года.

Изначально планировалось выпустить Btrfs v1.0 (и зафиксировать формат хранения на диске) в конце 2008 года, однако формат был зафиксирован только 12 июня 2010 года.

В файловой системе **ReiserFS** большое количество мелких файлов приводит к наибольшей «потере» дискового пространства [3]. Но при таком способе хранения данных переместить/создать/скопировать файл занимает гораздо меньше времени.

Преимуществом файловой системы ReFS является использование функций восстановления, встроенных в файловую систему, которые позволяют сразу исправлять ошибки, не запуская долгую глобальную проверку. Это прежде всего актуально для больших объемов, данных. ReFS может обрабатывать до 1ЙБ (Йоттабайт).

Для хранения информации о свободных объектах ReiserFS использует не простые списки, а несколько более сложные структуры данных. В системе ReiserFS для этого применяются так называемые "сбалансированные деревья" или "B+Trees", время поиска в которых пропорционально не количеству объектов (файлов в каталоге или числа блоков на диске), а логарифму этого числа. В сбалансированном дереве все ветви (пути от корня до "листа") имеют одинаковую (или примерно одинаковую) длину. ReiserFS использует сбалансированные деревья для хранения всех объектов файловой системы: файлов в каталогах, данных о свободных блоках и т. д. Это позволяет существенно повысить производительность обращения к дискам.

ReiserFS также имеет ряд особенностей, нацеленных специально для улучшения работы с маленькими файлами. ReiserFS не связана ограничением в ассигновании памяти для файла в целом числе 1-2-4 KB блоков. По необходимости для файла может ассигноваться точный размер. ReiserFS также включает некоторые виды специальной оптимизации файловых "хвостов" для хранения конечных частей файлов, меньших, чем логический блок файловой системы. Для увеличения скорости, ReiserFS способен хранить содержимое файлов непосредственно внутри дерева b*tree, а не в виде указателя на дисковый блок (в ext2 есть понятие fastlink, когда содержимое "мягкой" ссылки до 60 байт хранится в inode).

Тем самым достигается две вещи. Первое, сильно увеличивается производительность, так как данные и метаданные (stat_data, иначе говоря, inode) информация может хранится рядом и считываться одной дисковой операцией ввода/вывода. Второе, ReiserFS способен упаковать хвосты (tail) файлов, экономя дисковое пространство. Фактически, при разрешении ReiserFS выполнять упаковку хвостов (значение по умолчанию) будет экономиться примерно шесть процентов дискового пространства (в сравнении с ext2).

Следует иметь в виду, что упаковка хвостов требует дополнительной работы, так как при изменении размеров файлов необходима "переупаковка". По этой причине в ReiserFS упаковка хвоста может отключаться, позволяя администратору выбрать между скоростью и эффективностью использования дискового пространства.

Journaled File System или **JFS** – 64-битная журналируемая файловая система созданная IBM, доступная под лицензией GNU GPL.[1]

В операционной системе AIX существует два поколения JFS называемых JFS (JFS1) и JFS2 соответственно. В других операционных системах, таких как OS/2 и Linux,

существует только второе поколение, которое называется просто JFS. Также JFS называют файловую систему VxFS компании Veritas Software, используемую в ОС HP-UX.

Первоначально JFS была разработана корпорацией IBM для операционной системы AIX. JFS второго поколения была разработана IBM для ОС Warp Server for e-Business. Позже она была перенесена в IBMAIX и Linux. Целью разработчиков было обеспечить высокую производительность, надёжность и масштабируемость для многопроцессорных компьютеров.

В отличие от ext3, в которую поддержка журналирования была добавлена, JFS изначально была журналируемой. JFS ведет журнал только метаданных, поддерживая структуру файловой системы целостной, но не обязательно сохраняет данные. Отключение питания или крах системы может привести к сохранению устаревших копий файлов, однако сами файлы останутся пригодными к использованию. Журналирование JFS похоже на журналирование XFS, которая журналирует только части inode.

Для управления разделами диска в формате JFS был выпущен набор утилит под названием JFSutils.

3. Методы исследования

При тестировании скорости файловых систем был использованы:

- сервер на базе Core i7-4960 с 16Gb DDR3;
- жесткий диск (HDD) WD black 7000rpm;
- гипервизор VmWare ESXI 5.0;
- виртуальная машина с двумя ядрами и 8 Gb оперативной памяти;
- CENTOS 6 с последними обновлениями.

Для сравнения скоростей были использованы следующие сценарии Bash:

```
cmd1="cp -r /media/media4/video/best $dest"
cmd2="rsync -rlhtgopu /media/media4/backup $dest"
cmd3="grep linux -sir $dest/backup/wine-src/"
cmd4="find $dest -type f -delete"
```

Условия тестирования:

- виртуальная машина с 2 ядрами и 4 гб оперативной памяти;
- замеры проводились с помощью /usr/bin/time;
- между тестами 10 минутные паузы, чтобы устаканить uptime;
- размер раздела с ФС подобран так чтобы данные заполняли его на 2/3;
- размеры файлов применяемых в тесте.
- мелкие файлы – 1,7G /media/media1 - 40285 файлов.
- крупные файлв – 17,4G /media/media2 - 4 файла.

4. Результаты исследования

Сравнительные характеристики ФС приведены в таблице 2.

Таблица 2 – Сравнительные характеристики файловых систем

	ext2	**ext3**	**xfs**	**btrfs**	**reiserfs**	**jfs**
Содержимое папок	таблица	таблица	B+ деревья	B+ деревья	B+ деревья	B+ деревья
Размещение файлов	Битовая карта	Битовая карта	B+ деревья	экстент	Битовая карта	Битовая карта
Максимальный размер файла	От 16 Гигабайт до 2 Тирабайт	зависит от размера блока	8 Эксбибайт	8 Эксбибайт	1 Эксбибайт	4 Петабайта
Максимальная длина имени файла	255 байт	зависит от размера блока	256 байт	255 байт	4032 байт	255 байт
Максимальный размер тома	От 2 до 32 Тирабайт	зависит от размера блока	16 Эксбибайт	16 Эксбибайт	16 Терабайт	32 Петабайта

Сравнение производиельности файловых систем приведены в таблице 3 и на графиках, представленных рисунке 1 [16].

Таблица 3 - Сравнение производительности файловых систем

	ext2	**ext3**	**ext4**	**xfs**	**reiserfs**	**btrfs**	**jfs**
Копирование больших файлов:	116.03	122.69	116.45	137.47	138.67	130.25	130.98
Архивирование маленьких файлов	115.33	124.25	99.61	220.50	119.25	98.44	172.21
поиск среди маленьких файлов:	66.71	63.69	68.76	47.02	66.45	77.18	107.21
повторный поиск среди маленьких файлов:	100.47	97.27	102.36	80.70	96.48	101.27	135.29
поиск и удаление файлов:	8,09	7,51	6,40	82,59	10,22	13,53	15.67
средняя нагрузка на систему:	1.85, 1.37,	1.95, 1.39,	1.99, 1.26	2.02, 1.64	2.00, 1.47	2.09, 1.37	2.55, 1.99

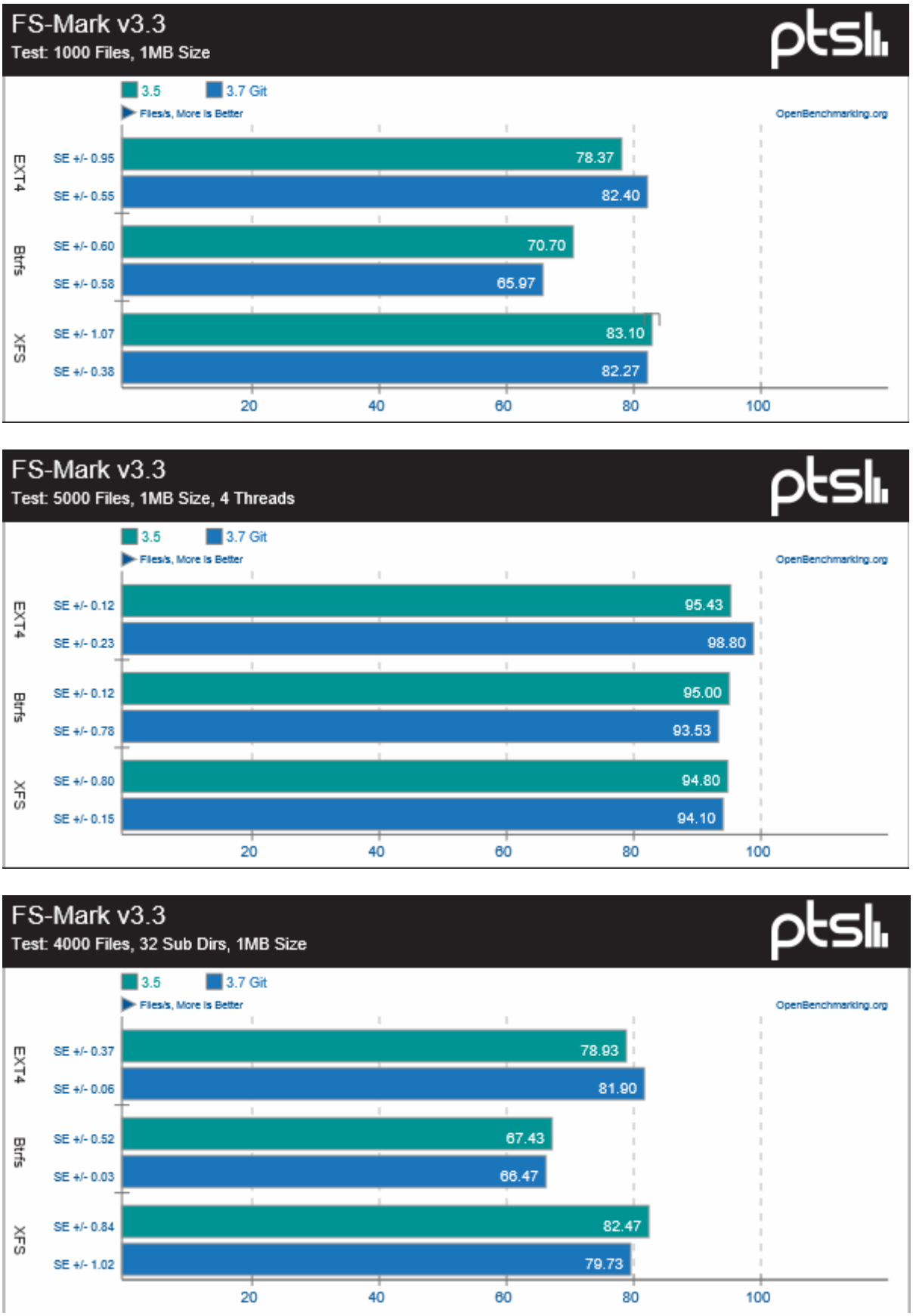

Рисунок 1 – Сравнение производиельности файловых систем ядра Linux

Исходя из данных таблицы 3, оптимальными по производительности являются файловые системы ext2 и ext, однако надо учитывать тот факт, что ext4 является журналируемой файловой системой и больше устойчива к сбоям.

5. Заключение

Рост нагрузок в локальных вычислительных сетях связан в большинстве случаев с увеличением файлоборота внутри локальной сети, что поднимает планку требований к файловым системам на серверах. В данной статье мы рассмотрели несколько файловых систем ядра Linux, активно использующихся в настоящее время. Характеристики файловых системы были приведены к структурированному виду и были проведены замеры скорости при различных операциях с файлами. Оптимальными показателями обладает файловая система ext4 используемая по-умолчанию во многих дистрибутивах Linux.

Список информационных источников

[1] Свободная энциклопедия: сайт «Википедия» [Электронный ресурс]: URL:https://ru.wikipedia.org/.
[2] https://ru.wikipedia.org/wiki/Ext4
[3] https://ru.wikipedia.org/wiki/ReiserFS
[4] https://ru.wikipedia.org/wiki/ExFAT
[5] http://habrahabr.ru/post/179821/
[6] http://citforum.ru/operating_systems/linux/robbins/fs02.shtml
[7] http://xgu.ru/wiki/ext4
[8] http://www.linuxcenter.ru/lib/books/kostromin/gl_16_08.phtml
[9] http://www.opennet.ru/docs/RUS/reiserfs_ondisk_layout/
[10] http://www.opennet.ru/opennews/art.shtml?num=21343
[11] http://habrahabr.ru/post/191136/
[12] http://habrahabr.ru/post/179821/
[13] http://habrahabr.ru/post/54043/
[14] http://habrahabr.ru/post/45873/
[15] http://www.linuxcenter.ru/lib/books/kostromin/gl_16_01.phtml
[16] http://www.phoronix.com/scan.php?page=article&item=linux_37_fsthree&num=2
[17] Остроух А.В. Ввод и обработка цифровой информации: учебник для нач. проф. образования / А.В. Остроух. – М.: Издательский центр «Академия», 2012. – 288 с. – ISBN 978-5-7695-9457-1.
[18] Остроух А.В. Основы информационных технологий: учебник для сред. проф. образования / А.В. Остроух. – М.: Издательский центр «Академия», 2014. – 208 с. – ISBN 978-5-4468-0588-4.
[19] Помазанов А.В., Остроух А.В. Создание и тестирование распределённой системы работы с удалёнными узлами // Автоматизация и современные технологии. – 2014. – №7. – С. 17-23.
[20] Сальный А.Г., Збавитель П.Ю., Николаев А.Б., Остроух А.В. Описание унифицированных программных модулей для лаборатории коллективного пользования // Автоматизация и управление в технических системах. – 2013. – № 2. – С. 12-17.