

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра математической
кибернетики и компьютерных наук

**ПОСТРОЕНИЕ МЕХАНИЗМА РЕЗЕРВНОГО КОПИРОВАНИЯ И
ВОССТАНОВЛЕНИЯ ФАЙЛОВ С УДАЛЕННЫМ ХРАНЕНИЕМ
КОПИЙ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Вакс Софьи Леонидовны

Научный руководитель
доцент, к. ф.-м. н.

В. Г. Самойлов

Заведующий кафедрой
к. ф.-м. н.

С. В. Миронов

Саратов 2016

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Организация систем резервного копирования и восстановления данных	4
1.1 Основные виды резервного копирования	4
1.2 Характеристики систем резервного копирования	5
1.3 Способы хранения резервных копий	5
1.4 Синхронизация	5
1.5 Файловая система Windows	6
1.6 Протоколы передачи данных	6
1.7 Анализ способов резервного копирования	7
2 Разработка программного обеспечения для резервного копирования и восстановления данных	8
2.1 Интерфейс и дополнительные возможности	8
2.2 Режим отслеживания и сохранения изменений	9
2.2.1 Механизм отслеживания изменений	9
2.2.2 Структура папки с историей изменений	9
2.2.3 Выбор отслеживаемой папки	10
2.3 Хранение копий на удаленной машине	10
2.3.1 SSH-соединение	10
2.3.2 Хранение инкрементных копий	10
2.3.3 Хранение полной копии	11
2.4 Режим синхронизации	11
2.5 Режим восстановления	12
2.5.1 Восстановление файла	12
2.5.2 Восстановление папки	12
2.6 Инструменты разработки	12
2.7 Сравнение программ для резервного копирования	12
ЗАКЛЮЧЕНИЕ	13
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	14

ВВЕДЕНИЕ

Разработка приложений для резервного копирования направлена на обеспечение безопасности данных при сбоях аппаратных средств, программного обеспечения, в непредвиденных ситуациях.

Актуальность темы дипломной работы объясняется наличием большого объема информации в электронном виде, которую необходимо защищать от потерь с помощью программ резервного копирования и восстановления данных.

Цель данной дипломной работы – разработать эффективный механизм резервного копирования и восстановления данных с удаленным хранением копий.

Для достижения поставленной цели дипломной работы необходимо решить следующие задачи:

- изучить основные принципы организации процесса резервного копирования и восстановления данных;
- выделить оптимальные способы резервного копирования, а также выбрать место хранения резервных копий и способ их переноса;
- разработать программное обеспечение, реализующее функции резервного копирования, синхронизации и восстановления данных;
- сравнить функциональность разработанного в ходе дипломной работы программного обеспечения и уже существующих продуктов.

Структурно дипломная работа состоит из введения, двух глав и заключения.

В первой главе рассматриваются теоретические основы организации систем резервного копирования, а также определяется оптимальная стратегия для разработки программного обеспечения.

Во второй главе рассматриваются этапы разработки программного обеспечения, реализующего механизм резервного копирования и восстановления данных. Также приводится сравнение функциональных возможностей существующих программ для резервного копирования и программного обеспечения, разработанного в ходе дипломной работы.

1 Организация систем резервного копирования и восстановления данных

Рассмотрим определение понятия резервного копирования по версии SNIA (Storage Networking Industry Association) [1]. Резервная копия (англ. backup copy) – данные, хранимые на энергонезависимых носителях, обычно удаленно, предназначенные для восстановления, в случае если оригинал копии данных утерян или недоступен. Резервное копирование (англ. backup) – процесс создания резервных копий.

Существует несколько распространенных операций системного копирования:

- полное резервное копирование;
- инкрементное резервное копирование;
- дифференциальное резервное копирование.

1.1 Основные виды резервного копирования

Полная резервная копия содержит все данные о папках и файлах, которые выбраны для резервного копирования, поэтому их быстрее и проще восстановить. [2]

Одна полная резервная копия обеспечивает возможность полностью восстановить все требуемые файлы и папки. [3]

Полное резервное копирование часто ограничивается еженедельным или ежемесячным графиком, хотя при достаточно большой скорости и емкости резервного носителя появляется возможность делать полное резервное копирование каждую ночь.

При инкрементном резервном копировании хранятся резервные копии всех файлов, которые изменились с момента последнего полного, дифференциального или инкрементного резервного копирования. [2]

Преимуществом инкрементного резервного копирования является минимальное время его выполнения.

Процесс восстановления информации после инкрементного резервного копирования занимает большее количество времени в сравнении с иными видами резервного копирования. При восстановлении необходима полная резервная копия, а также каждая инкрементная резервная копия, которая была создана с момента последнего полного резервного копирования.

Дифференциальная резервная копия содержит все файлы, которые были изменены с момента последнего полного резервного копирования. [2]

Восстановление дифференциальной резервной копии происходит быстрее, чем процесс восстановления инкрементных резервных копий, поскольку необходимы только две резервные копии: последняя полная резервная копия и последняя дифференциальная. [3]

Если несколько раз запускать дифференциальное резервное копирование после полного резервного копирования, создадутся копии тех файлов, которые уже были включены в предыдущие дифференциальные резервные копии и не были изменены. При планировании расписания нужно учитывать время на создание последней самой большой дифференциальной копии [4].

1.2 Характеристики систем резервного копирования

Важную роль при выборе систем резервного копирования играют характеристики процессов хранения и резервирования. Их необходимо оговаривать в контракте с поставщиком программного обеспечения. Требования к данным характеристикам описываются в отдельном соглашении об уровне услуг SLA (англ. Service Level Agreement).

При выдвижении требований к характеристикам обычно оперируют такими терминами, как RPO, RTO, Backup Window. [5]

1.3 Способы хранения резервных копий

Основные места хранения резервных копий:

- жесткий диск;
- магнитная лента;
- твердотельный накопитель;
- оптические диски;
- удаленное сетевое хранилище.

1.4 Синхронизация

Синхронизация – процесс согласования или приведения двух и более объектов к эквивалентному состоянию в заданный момент времени.

При односторонней синхронизации файлы будут изменяться только в одном расположении. [6]

При двусторонней синхронизации файлы копируются в двух направлениях, благодаря чему в обоих местах сохраняются идентичные файлы. Добавление, изменение или удаление файла в одном расположении приводит к таким же изменениям в другом расположении.

1.5 Файловая система Windows

Операционная система Windows поддерживает несколько файловых систем: NTFS, FAT и FAT32. Но работать может только на NTFS, то есть установлена может быть только на раздел жесткого диска, отформатированного в данной файловой системе. Обусловлено это теми особенностями и инструментами безопасности, которые предусмотрены в NTFS, но отсутствуют в файловых системах Windows предыдущего поколения: FAT16 и FAT32. [7]

В зависимости от файловой системы, файл может обладать различным набором свойств [8]:

- имя;
- расширение;
- основные атрибуты;
- время;
- владелец и группа файлов;
- права доступа.

1.6 Протоколы передачи данных

Протокол SSH (Secure Shell, Безопасная оболочка) охватывает аутентификацию, шифрование и целостность данных, передаваемых по сети. SSH использует современные, безопасные алгоритмы шифрования, которые достаточно эффективны, чтобы использоваться в критически важных приложениях в крупных корпорациях.

Протокол FTP (File Transfer Protocol, Протокол передачи файлов) отвечает за пересылку большинства файлов по Интернету; этот протокол позволяет как удаленным, так и локальным клиентам или серверам осуществлять эффективную передачу файлов или данных, используя при этом надежный транспортный протокол TCP.

1. FTP является протоколом передачи файлов, в то время как SSH это сетевой протокол.

2. FTP является небезопасным протоколом передачи данных, в то время как SSH по своей сути обеспечивает необходимую безопасность.
3. FTP позволяет управлять только файлами на удаленной машине, в то время как у SSH более широкий спектр применения.

1.7 Анализ способов резервного копирования

В случае, если приоритетом при разработке программного обеспечения является быстрота создания резервных копий и экономия дискового пространства, необходимо использовать инкрементное резервное копирования. При использовании данного вида резервного копирования необходимо изначально создавать полную резервную копию. Для быстрого восстановления данных следует использовать полное резервное копирование. Также синхронизация является оптимальным способом резервного копирования. В случае, если данные потеряны на одной из машин и присутствуют на другой, с помощью синхронизации можно перенести данные на пустую машину.

Оптимальным выбором для хранения резервных копий является удаленное сетевое хранилище. Такой способ позволяет сэкономить место на локальной машине, получить доступ из любой точки, где подключен Интернет, а также получить большой объем свободной памяти на удаленной машине. Осуществлять передачу данных на удаленную машину лучше по протоколу SSH, который обеспечивает шифрование передаваемых данных.

Таким образом, разработанное в данной дипломной работе программное обеспечение базируется на указанных выше выводах.

2 Разработка программного обеспечения для резервного копирования и восстановления данных

При разработке программного обеспечения были поставлены следующие требования:

1. Разработать удобный для пользователя интерфейс.
2. Реализовать возможность выбора пользователем папки, включающей в себя вложенные папки и файлы любого типа, для резервного копирования и дальнейшего восстановления.
3. Организовать отслеживание изменений в папке, которая является хранищем файлов, фиксируя произошедшие изменения для каждого файла, создавая их копии.
4. Устаревшие копии файлов, а также полную резервную копию папки, хранить на удаленной машине.
5. Обеспечить возможность двусторонней синхронизации между локальной и удаленной машинами.
6. Реализовать восстановление папки на любой момент времени.
7. Реализовать восстановление отдельного файла на любой момент времени.
8. Реализовать программу на языке программирования C# под операционную систему Windows.

2.1 Интерфейс и дополнительные возможности

User Friendly интерфейс программного обеспечения разработан с помощью приложения Windows Forms. В окне программы присутствует четыре вкладки: «Status», «Backup», «Synchronization», «Settings».

После запуска значок программы можно увидеть в области уведомлений. При нажатии правой кнопки мыши на значок появляется контекстное меню.

При выборе пункта «Show or hide» приложение сворачивается, если окно программы было видимым, или разворачивается, если было скрыто.

При выборе пункта «Open local dir» открывается указанная отслеживаемая папка.

При выборе пункта «Exit» приложение закрывается.

2.2 Режим отслеживания и сохранения изменений

2.2.1 Механизм отслеживания изменений

При фиксировании изменений в реальном времени применяется инкрементное резервное копирование. Создаются единичные копии файлов, которые подверглись изменению.

Но для корректной работы с инкрементными копиями необходимо изначально создать полную резервную копию данных. Полная резервная копия создается тогда, когда это первый запуск программы и отслеживаемая папка не пуста. Если программа запущена не в первый раз, то на локальной машине уже хранится структура с полной резервной копией данных.

Режим наблюдения в операционной системе Windows осуществляется с помощью класса `FileSystemWatcher`. Данный класс ожидает уведомления файловой системы об изменениях и инициирует события (`Changed`, `Renamed`, `Deleted`, `Created`) при изменениях папки или файла в папке.

Программа отслеживает все виды изменений:

- удаление;
- создание;
- переименование;
- изменение содержимого.

Изменения отображаются в окне программы.

2.2.2 Структура папки с историей изменений

При каждом изменении в отслеживаемой папке данное событие фиксируется в истории изменений, которая хранится в папке сложенными папками, `log` файлами и копиями файлов.

Для каждой папки или файла генерируется уникальный 128-битный идентификатор `GUID` [9], который в папке `copies` становится именем папки, содержащей `log`-файл и копии файла.

В случае возникновения события `Changed`, которое возникает только при изменении содержимого файла, создается копия, в названии которой записывается дата произошедшего изменения.

Для каждого файла или папки создается `log`-файл, в котором фиксируются произошедшие изменения с конкретным объектом.

2.2.3 Выбор отслеживаемой папки

Выбор отслеживаемой папки осуществляется пользователем во вкладке «Settings» в поле «Monitored folder». Для выбора новой отслеживаемой папки необходимо ввести путь к папке и нажать кнопку «Save settings».

Для каждой отслеживаемой папки присваивается свой идентификатор GUID. Создается папка с названием присвоенного GUID. Каждая такая папка содержит:

- папку copies с историей изменений;
- файл dates.txt с датами всех произошедших изменений;
- файл guidFiles.txt с путями к файлам и присвоенными им GUID;
- файл guidDirectories.txt с путями папок и присвоенными им GUID;
- файл lastdate.txt с датой последнего полного резервного копирования на удаленную машину.

2.3 Хранение копий на удаленной машине

2.3.1 SSH-соединение

Для передачи данных на удаленную машину был выбран протокол SSH, так как передаваемые данные шифруются и обеспечивается большая конфиденциальность данных.

В качестве удаленной машины использована виртуальная машина VirtualBox с установленной операционной системой RedHat Enterprise Linux 5.

На удаленную машину необходима установка пакета openssh для корректного соединения по протоколу SSH.

В программе есть возможность установить настройки соединения по SSH для удаленной машины.

2.3.2 Хранение инкрементных копий

В том случае, если копии файлов хранятся на локальной машине больше недели, данные копии переносятся на удаленную машину.

PuTTY – свободно распространяемый клиент для различных протоколов удалённого доступа. Для выполнения данных задач используются утилиты, входящие в пакет PuTTY:

- создать папку с заданным именем можно с помощью утилиты plink, которая является консольной заменой приложения PuTTY;

- перенос копий файлов по SSH осуществляется с помощью утилиты pscp, являющейся клиентом для SCP (удаленное копирование файлов по шифрованному протоколу scp с управлением из командной строки).

2.3.3 Хранение полной копии

На случай полной потери данных на локальной машине на удаленную машину переносится полная резервная копия папки.

Полная резервная копия должна быть не старше недели, поэтому дата переноса полной резервной копии записывается в файл lastdate.txt.

Перенос полной резервной копии осуществляется следующим образом. Сначала на локальной машине создается архив формата zip, в котором хранится вся структура отслеживаемой папки. Далее данный архив копируется на удаленную машину с помощью директивы pscp, после чего удаляется с локальной.

2.4 Режим синхронизации

В данной дипломной работе реализован механизм двусторонней синхронизации данных между локальной и удаленной машинами. Это еще один механизм резервного копирования и восстановления данных.

Если на удаленной машине будут обновления в синхронизируемой папке, то эти изменения отразятся на отслеживаемой папке локальной машины. В процессе синхронизации применяется инкрементное резервное копирование данных. [6]

Синхронизация происходит с той папкой, которую укажет пользователь в специальном окне. Эта папка будет вложена в основную удаленную папку, указанную в окне «Settings». Процесс синхронизации происходит принудительно при нажатии кнопки «Start» по желанию пользователя.

Для обеспечения двусторонней синхронизации папок используется свободно распространяемое программное обеспечение Unison. Передача данных осуществляется по протоколу SSH.

Для корректной работы приложения на удаленную машину необходимо установить пакет unison.

2.5 Режим восстановления

2.5.1 Восстановление файла

В программе пользователю предлагается выбрать для восстановления один из файлов, который когда-либо присутствовал в отслеживаемой папке, и одну из дат изменений этого файла. Для начала восстановления необходимо нажать кнопку «Start BackUp».

2.5.2 Восстановление папки

Если требуется полностью восстановить папку на определенное время со всеми вложенными файлами и папками, то в программе предлагается ввести дату, к которой пользователь хочет откатить отслеживаемую папку. Для начала восстановления необходимо нажать кнопку «Start BackUp».

2.6 Инструменты разработки

В ходе выполнения дипломной работы была использована ЭВМ с процессором с тактовой частотой 2,3 ГГц и 4 ГБ ОЗУ. Операционная система Windows 10 и среда разработки VisualStudio Professional 2013. Язык программирования, используемый для реализации программы – C#.

Для корректной работы программы на используемой ОС Windows должен быть установлен .Net Framework. Во многих версиях Windows .Net Framework включен по умолчанию.

2.7 Сравнение программ для резервного копирования

Было проведено сравнение различных программ для резервного копирования информации. Основными критериями для сравнения выбраны способ резервного копирования (инкрементное, полное), место хранения копий, возможность синхронизации, возможность сжатия копий, возможность работы по расписанию, совместимость с различными операционными системами.

ЗАКЛЮЧЕНИЕ

В данной работе описаны этапы разработки собственного программного обеспечения, реализующего механизм резервного копирования, синхронизации и восстановления данных.

Реализованы возможности отслеживания изменений в папке, восстановления отдельного файла или целой папки на указанный момент времени, синхронизации удаленной и локальной папки. Те резервные копии, которым больше недели на момент запуска программы, хранятся на удаленной машине. Также раз в неделю обновляется полная резервная копия, сжатая в zip-архив. Кроме того, создан удобный для пользователей интерфейс.

Основными преимуществами данной программы являются экономия дискового пространства на локальной машине за счет сохранения устаревших инкрементных копий на удаленной машине и экономия дискового пространства на удаленной машине за счет сжатия полной резервной копии. Безопасность передачи данных на удаленную машину обеспечена с помощью протокола безопасного соединения SSH.

Кроме того, организовано быстрое восстановление на недавний период времени за счет сохранения новых инкрементных копий на локальной машине. Также реализована синхронизация данных с удаленной машиной, что позволяет быстро восстанавливать данные. Существует возможность расширения функционала под поставленные задачи.

Разработанное программное обеспечение является готовым продуктом, которое может использоваться как рядовыми пользователями, так и коммерческими организациями в целях сохранения и защиты информации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 A Dictionary of Storage Networking Terminology 2016 [Электронный ресурс]. — URL: <http://www.snia.org/education/dictionary> (Дата обращения 20.02.2016). Загл. с экрана. Яз. англ.
- 2 Казаков, В. Г. Моделирование алгоритмов резервного копирования для получения оценок объема репозитория / В. Г. Казаков, С. А. Федосин // *Информационные технологии*. — 2008. — Vol. 6, no. 3. — Pp. 126–132.
- 3 Berman, J. Backup. Acronis Special Edition / J. Berman. — Hoboken, USA: Wiley, 2014.
- 4 Dorion, P. Backup and Disaster Recovery. Incremental/differential backup differences. [Электронный ресурс] / P. Dorion. — 2005. — URL: http://searchstorage.techtarget.com/tip/1,289483,sid5_gci1119483,00.html (Дата обращения: 25.04.2016) Загл. с экрана. Яз. англ.
- 5 Coughlin, T. Data Protection for Small Businesses and Homes / T. Coughlin. — Coughlin Associates, Inc., 2007.
- 6 Tridgell, A. Efficient Algorithms for Sorting and Synchronization: Ph.D. thesis / The Australian National University. — 1999.
- 7 Касперски, К. Восстановление данных. Практическое руководство / К. Касперски. — СПб.: БХВ-Петербург, 2007.
- 8 Обзор файловых систем fat, hpfs и ntfs [электронный ресурс]. — URL: <https://support.microsoft.com/ru-ru/kb/100108/ru> (дата обращения: 30.04.2016).
- 9 GUID structure [Электронный ресурс]. — URL: [https://msdn.microsoft.com/en-us/library/aa373931\(VS.85\).aspx](https://msdn.microsoft.com/en-us/library/aa373931(VS.85).aspx) (Дата обращения 20.02.2016). Загл. с экрана. Яз. англ.