

wget

В этой статье рассмотрим утилиту `wget`, которая позволяет загружать файлы из Всемирной паутины (WWW) с использованием широко распространенных протоколов, таких как HTTP, HTTPS, FTP и FTPS.

`Wget` - это бесплатная утилита командной строки и сетевой загрузчик файлов, которая обладает множеством функций, облегчающих загрузку файлов, включая:

- Загрузка больших файлов или зеркальное копирование веб- или FTP-сайтов.
- Загрузка нескольких файлов одновременно.
- Установка ограничения на пропускную способность и скорость загрузки.
- Загрузка файлов через прокси-серверы.
- Возможность возобновления прерванных загрузок.
- Рекурсивное зеркалирование каталогов.
- Работа на большинстве UNIX-подобных операционных систем, а также на Windows.
- Не требует внимания пользователя / фоновая работа.
- Поддержка постоянных HTTP-соединений.
- Поддержка SSL/TLS для зашифрованных загрузок с использованием библиотек OpenSSL или GnuTLS.
- Поддержка загрузок по IPv4 и IPv6.

Синтаксис команды Wget

Базовый синтаксис `Wget` выглядит следующим образом:

```
$ wget [опция] [URL]
```

Сначала проверьте, установлена ли утилита `wget` на вашем компьютере с Linux, используя следующую команду.

```
$ rpm -q wget [для RHEL/CentOS/Fedora и Rocky Linux/AlmaLinux]
$ dpkg -l | grep wget [для Debian, Ubuntu и Mint]
```

Установка Wget на Linux

Если `Wget` не установлен, вы можете установить его с помощью стандартного менеджера пакетов вашей системы Linux, как показано ниже.

```
$ sudo apt install wget -y [для Debian, Ubuntu и Mint]
$ sudo yum install wget -y [для RHEL/CentOS/Fedora и Rocky Linux/AlmaLinux]
$ sudo emerge -a net-misc/wget [для Gentoo Linux]
$ sudo pacman -Sy wget [для Arch Linux]
$ sudo zypper install wget [для OpenSUSE]
```

Опция **-y**, использованная здесь, предотвращает запросы на подтверждение перед установкой любого пакета.

Загрузка файла с помощью Wget

Команда загружает один файл и сохраняет его в текущем каталоге. Также она отображает ход загрузки, размер, дату и время в процессе.

```
wget http://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz

--2021-12-10 04:15:16--  http://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz
Resolving ftp.gnu.org (ftp.gnu.org)... 209.51.188.20, 2001:470:142:3::b
Connecting to ftp.gnu.org (ftp.gnu.org)|209.51.188.20|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3565643 (3.4M) [application/x-gzip]
Saving to: 'wget2-2.0.0.tar.gz'

wget2-2.0.0.tar.gz 100%[=====>] 3.40M  2.31MB/s in 1.5s

2021-12-10 04:15:18 (2.31 MB/s) - 'wget2-2.0.0.tar.gz' saved [3565643/3565643]
```

Загрузка файла с помощью Wget под другим именем

Используя опцию **-O** (прописная буква), можно загрузить файлы с другими именами. В примере ниже указано имя файла как wget.zip.

```
wget -O wget.zip http://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz

--2021-12-10 04:20:19--  http://ftp.gnu.org/gnu/wget/wget-1.5.3.tar.gz
Resolving ftp.gnu.org (ftp.gnu.org)... 209.51.188.20, 2001:470:142:3::b
Connecting to ftp.gnu.org (ftp.gnu.org)|209.51.188.20|:80... connected.
```

```
HTTP request sent, awaiting response... 200 OK
Length: 446966 (436K) [application/x-gzip]
Saving to: 'wget.zip'

wget.zip 100%[=====>] 436.49K 510KB/s in 0.9s

2021-12-10 04:20:21 (510 KB/s) - 'wget.zip' saved [446966/446966]
```

Загрузка нескольких файлов с помощью протоколов HTTP и FTP через Wget

Давайте рассмотрим, как одновременно загрузить несколько файлов с использованием протоколов HTTP и FTP с помощью команды wget.

```
wget http://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz ftp://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz.sig

--2021-12-10 06:45:17-- http://ftp.gnu.org/gnu/wget/wget2-2.0.0.tar.gz
Resolving ftp.gnu.org (ftp.gnu.org)... 209.51.188.20, 2001:470:142:3::b
Connecting to ftp.gnu.org (ftp.gnu.org)|209.51.188.20|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3565643 (3.4M) [application/x-gzip]
Saving to: 'wget2-2.0.0.tar.gz'

wget2-2.0.0.tar.gz 100%[=====>] 4.40M 4.31MB/s in 1.1s

2021-12-10 06:46:10 (2.31 MB/s) - 'wget2-2.0.0.tar.gz' saved [3565643/3565643]
```

Загрузка нескольких файлов из файла с помощью Wget

Чтобы загрузить несколько файлов одновременно, используйте опцию -i с указанием расположения файла, который содержит список URL-адресов для загрузки. Каждый URL-адрес должен быть добавлен на отдельной строке.

Например, следующий файл «download-linux.txt» содержит список URL-адресов для загрузки.

```
cat download-linux.txt
```

```
https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
https://download.rockylinux.org/pub/rocky/8/isos/x86_64/Rocky-8.5-x86_64-dvd1.iso
https://cdimage.debian.org/debian-cd/current/amd64/iso-dvd/debian-11.2.0-amd64-DVD-1.iso
```

```
wget -i download-linux.txt
```

```
--2021-12-10 04:52:40-- https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
Resolving releases.ubuntu.com (releases.ubuntu.com)... 91.189.88.248, 91.189.88.247,
91.189.91.124, ...
Connecting to releases.ubuntu.com (releases.ubuntu.com)|91.189.88.248|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3071934464 (2.9G) [application/x-iso9660-image]
Saving to: 'ubuntu-20.04.3-desktop-amd64.iso'

ubuntu-20.04.3-desktop-amd64 4%[=> ] 137.71M 11.2MB/s eta 3m 30s
...
```

Если список ваших URL-адресов имеет определенный шаблон нумерации, вы можете использовать фигурные скобки, чтобы загрузить все URL-адреса, соответствующие шаблону. Например, если вы хотите загрузить серию ядер Linux, начиная с версии 5.1.1 до 5.1.15, вы можете сделать следующее.

```
wget https://mirrors.edge.kernel.org/pub/linux/kernel/v5.x/linux-5.1.{1..15}.tar.gz
```

```
--2021-12-10 05:46:59-- https://mirrors.edge.kernel.org/pub/linux/kernel/v5.x/linux-5.1.1.tar.gz
Resolving mirrors.edge.kernel.org (mirrors.edge.kernel.org)... 147.75.95.133,
2604:1380:3000:1500::1
Connecting to mirrors.edge.kernel.org (mirrors.edge.kernel.org)|147.75.95.133|:443...
connected.
WARNING: The certificate of 'mirrors.edge.kernel.org' is not trusted.
WARNING: The certificate of 'mirrors.edge.kernel.org' is not yet activated.
The certificate has not yet been activated
HTTP request sent, awaiting response... 200 OK
Length: 164113671 (157M) [application/x-gzip]
Saving to: 'linux-5.1.1.tar.gz'
```

```
linux-5.1.1.tar.gz 100%[=====>] 156.51M 2.59MB/s in 61s

2021-12-10 05:48:01 (2.57 MB/s) - 'linux-5.1.1.tar.gz' saved [164113671/164113671]

--2021-12-10 05:48:01-- https://mirrors.edge.kernel.org/pub/linux/kernel/v5.x/linux-5.1.2.tar.gz
Reusing existing connection to mirrors.edge.kernel.org: 443.
HTTP request sent, awaiting response... 200 OK
Length: 164110470 (157M) [application/x-gzip]
Saving to: 'linux-5.1.2.tar.gz'

linux-5.1.2.tar.gz 19%[=====] 30.57M 2.58MB/s eta 50s
```

Возобновление незавершенной загрузки с помощью Wget

При загрузке больших файлов иногда может случиться, что загрузка прерывается. В таких случаях можно возобновить загрузку того же файла с того места, на котором она была прервана, используя опцию `-c`.

Однако, когда вы начинаете загружать файлы без указания опции `-c`, `wget` добавит расширение `.1` в конец имени файла, и это будет считаться новой загрузкой. Поэтому, когда вы загружаете большие файлы, рекомендуется добавлять опцию `-c`.

```
wget -c https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso

--2021-12-10 05:27:59-- https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
Resolving releases.ubuntu.com (releases.ubuntu.com)... 91.189.88.247, 91.189.91.123,
91.189.91.124, ...
Connecting to releases.ubuntu.com (releases.ubuntu.com)|91.189.88.247|: 443... connected.
HTTP request sent, awaiting response... 206 Partial Content
Length: 3071934464 (2.9G), 2922987520 (2.7G) remaining [application/x-iso9660-image]
Saving to: 'ubuntu-20.04.3-desktop-amd64.iso'

ubuntu-20.04.3-desktop-amd64.iso 5%[++++++> ] 167.93M 11.1MB/s
^C
[[email protected] ~]# wget -c https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-
amd64.iso
```

```
--2021-12-10 05:28:03-- https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
Resolving releases.ubuntu.com (releases.ubuntu.com)... 91.189.88.248, 91.189.91.124,
91.189.91.123, ...
Connecting to releases.ubuntu.com (releases.ubuntu.com)|91.189.88.248|:443... connected.
HTTP request sent, awaiting response... 206 Partial Content
Length: 3071934464 (2.9G), 2894266368 (2.7G) remaining [application/x-iso9660-image]
Saving to: 'ubuntu-20.04.3-desktop-amd64.iso'

ubuntu-20.04.3-desktop-amd64.iso 10%[++++++=====> ] 296.32M 17.2MB/s eta 2m 49s ^
```

Зеркалирование (копирование) всего сайта с помощью Wget

Чтобы загрузить, создать зеркало или копировать весь сайт для просмотра в автономном режиме, можно использовать следующую команду, которая создаст локальную копию сайта вместе со всеми ресурсами (JavaScript, CSS, изображения).

```
wget --recursive --page-requisites --adjust-extension --span-hosts --convert-links --restrict-file-names=windows --domains yoursite.com --no-parent yoursite.com
```

Разберем опции команды:

- `--recursive`: Загружает весь сайт рекурсивно.
- `--page-requisites`: Получает все элементы/ресурсы страницы (CSS/JS/изображения).
- `--adjust-extension`: Сохраняет файлы с расширением ".html" на конце.
- `--span-hosts`: Включает необходимые ресурсы с других сайтов.
- `--convert-links`: Обновляет ссылки для работы в статической версии сайта.
- `--restrict-file-names=windows`: Изменяет имена файлов, чтобы они работали также в Windows.
- `--domains yoursite.com`: Не следует по ссылкам вне этого домена.
- `--no-parent`: Не следует по ссылкам вне указанного каталога.
- `yoursite.com`: URL-адрес для загрузки.

Эта команда создает локальную копию сайта, которую можно просматривать в автономном режиме, с корректными ссылками и необходимыми ресурсами.

Загрузка файлов в фоновом режиме с помощью Wget

С опцией `-b` вы можете отправить загрузку в фоновый режим сразу после начала загрузки, и логи будут записаны в файл `wget.log`.

```
wget -b wget.log https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
```

```
Continuing in background, pid 8999.
```

```
Output will be written to 'wget.log'.
```

Ограничение скорости загрузки файлов с помощью Wget

С опцией `--limit-rate=100k` скорость загрузки ограничивается до 100 кбит/с, и логи будут созданы в файле `wget.log`, как показано ниже.

```
wget -c --limit-rate=100k -b wget.log https://releases.ubuntu.com/20.04.3/ubuntu-20.04.3-desktop-amd64.iso
```

```
Continuing in background, pid 9108.
```

```
Output will be written to 'wget-log'.
```

Для просмотра содержимого файла `wget.log` и проверки скорости загрузки с использованием `wget` выполните следующую команду:

```
tail -f wget-log
```

5600K	0%	104K	8h19m
5650K	0%	103K	8h19m
5700K	0%	105K	8h19m
5750K	0%	104K	8h18m
5800K	0%	104K	8h18m
5850K	0%	105K	8h18m
5900K	0%	103K	8h18m
5950K	0%	105K	8h18m
6000K	0%	69.0K	8h20m
6050K	0%	106K	8h19m
6100K	0%	98.5K	8h20m
6150K	0%	110K	8h19m
6200K	0%	104K	8h19m
6250K	0%	104K	8h19m

...

Эта команда отобразит содержимое файла `wget.log` в терминале. В логах вы сможете увидеть различную информацию о загрузке, включая скорость загрузки. Обратите внимание, что скорость загрузки может колебаться в зависимости от вашего интернет-соединения и местоположения сервера.

Загрузка файлов, защищенных паролем, через FTP и HTTP с использованием Wget

Для загрузки файла с защищенного паролем FTP-сервера используйте опции `--ftp-user=username` и `--ftp-password=password`, как показано ниже.

```
$ wget --ftp-user=narad --ftp-password=password ftp://ftp.example.com/filename.tar.gz
```

Для загрузки файла с защищенного паролем HTTP-сервера используйте опции `--http-user=username` и `--http-password=password`, как показано ниже.

```
$ wget --http-user=narad --http-password=password http://http.example.com/filename.tar.gz
```

Замените "username" и "password" на соответствующие имя пользователя и пароль для доступа к защищенным ресурсам на FTP или HTTP сервере.

Игнорирование проверки SSL-сертификата с помощью Wget

Чтобы игнорировать проверку SSL-сертификата при загрузке файлов через HTTPS, используйте опцию `--no-check-certificate`:

```
wget --no-check-certificate https://mirrors.edge.kernel.org/pub/linux/kernel/v5.x/linux-5.1.1.tar.gz
```

```
--2021-12-10 06:21:21-- https://mirrors.edge.kernel.org/pub/linux/kernel/v5.x/linux-5.1.1.tar.gz
```

```
Resolving mirrors.edge.kernel.org (mirrors.edge.kernel.org)... 147.75.95.133,  
2604:1380:3000:1500::1
```



```
Connecting to mirrors.edge.kernel.org (mirrors.edge.kernel.org)| 147. 75. 95. 133| : 443...
connected.
WARNING: The certificate of 'mirrors.edge.kernel.org' is not trusted.
WARNING: The certificate of 'mirrors.edge.kernel.org' is not yet activated.
The certificate has not yet been activated
HTTP request sent, awaiting response... 200 OK
Length: 164113671 (157M) [application/x-gzip]
Saving to: 'linux-5.1.1.tar.gz'
...
```

Обратите внимание, что использование этой опции может представлять определенные риски для безопасности, так как она отключает проверку SSL-сертификатов и делает вас уязвимым для атак "человек посередине" (Man-in-the-Middle). Используйте эту опцию только в тех случаях, когда вы уверены в надежности источника или для тестирования в безопасной среде.

Дополнительные опции

- `--tries=number` - Устанавливает количество попыток для загрузки файла (по умолчанию 20). Если вы хотите изменить количество попыток, замените "number" на желаемое число.

```
wget --tries=5 https://example.com/filename.tar.gz
```

- `--timeout=seconds` - Устанавливает таймаут для операций чтения и соединения. Замените "seconds" на желаемое количество секунд.

```
wget --timeout=10 https://example.com/filename.tar.gz
```

- `--wait=seconds` - Ожидает указанное количество секунд между загрузками. Может быть полезно при обращении к серверам с ограничениями на количество запросов в единицу времени.

```
wget --wait=5 https://example.com/filename.tar.gz
```

- `--retry-connrefused` - Повторяет попытку соединения, даже если сервер отказал в соединении. Это может быть полезно при временных проблемах с доступом к серверу.

```
wget --retry-connrefused https://example.com/filename.tar.gz
```

- `--quiet` - Отключает вывод всех сообщений, кроме ошибок.

```
wget --quiet https://example.com/filename.tar.gz
```

- `--timestamping` - Загружает файлы только в том случае, если они были изменены на сервере по сравнению с локальной версией.

```
wget --timestamping https://example.com/filename.tar.gz
```

- `--no-cookies` - Отключает использование куки.

```
wget --no-cookies https://example.com/filename.tar.gz
```

- `--header="Header-Name: value"` - Добавляет пользовательский заголовок HTTP в запрос.

```
wget --header="User-Agent: Mozilla/5.0" https://example.com/filename.tar.gz
```

Это лишь некоторые из множества опций, доступных в Wget. Вся информация о доступных опциях может быть найдена в справке команды `wget --help` или в документации программы.

Revision #2

Created 2 May 2023 21:44:49 by Zerr0

Updated 2 May 2023 22:35:19 by Zerr0