

gzip

Da Wikipedia, l'enciclopedia libera.

Questa voce o sezione sugli argomenti software libero e formati di file non cita le fonti necessarie o quelle presenti sono insufficienti.

gzip (o **GNU Gzip**^[1]) è un software libero per la compressione dei dati, distribuito dalla Free Software Foundation^[2]. Creato da Jean-Loup Gailly e Mark Adler^[3] per sostituire l'utility compress^[4] a causa del brevetto sull'algoritmo LZW^[5], fu pubblicato inizialmente il 31 ottobre 1992, nella versione 0.1. La versione 1.0 fu distribuita solamente a partire dal febbraio 1993.

Nei sistemi operativi FreeBSD, NetBSD e OpenBSD è implementata una versione di gzip che è un *front-end* sulla libreria zlib.^{[6][7]}

Indice

Descrizione

Formato del file

Altri usi

Note

Bibliografia

Voci correlate

Altri progetti

Collegamenti esterni

Descrizione

Normalmente ogni archivio verrà rimpiazzato da uno con l'estensione .gz, mantenendo le stesse proprietà, date d'accesso e di modifica (l'estensione predefinita è gz per Linux o OpenVMS, z per MS-DOS, OS/2 FAT, Windows NT FAT e Atari). Qualora il nome file compresso fosse troppo lungo, esso verrà troncato.

Se non è specificato alcun file, o se un nome di file è "-", lo standard input è compresso nello standard output. Gzip proverà a comprimere solo i file regolari. In particolare

gzip software



```
!:/tmp/pdf2ps:$gzip -h
gzip 1.3.5
(2002-09-30)
usage: gzip [-cdfhlnNrtvV19] [-S suffix] [file ...]
-c --stdout      write on standard output, keep orig
-d --decompress  decompress
-f --force       force overwrite of output file and
-h --help       give this help
-l --list        list compressed file contents
-L --license     display software license
-n --no-name     do not save or restore the original
```

Genere	Compressione dei dati
Sviluppatore	progetto GNU
Data prima versione	1992
Ultima versione	1.12 (7 aprile 2022)
Sistema operativo	Unix-like
Linguaggio	C
Licenza	GNU GPL v3+ (licenza libera)
Sito web	www.gnu.org/software/gzip/ (http s://www.gnu.org/software/gzip/)

gzip	
Estensione	.gz
Tipo MIME	application/gzip
Uniform Type Identifier (UTI)	org.gnu.gnu-zip-archive

ignorerà i link simbolici.

Se il nome del file compresso è troppo lungo per il file system, gzip lo troncherà. Gzip prova a troncare solo le parti del nome del file più lunghe di 3 caratteri (una parte è delimitata da puntini). Se il nome consiste solo di parti piccole, le più lunghe sono troncate. Per esempio, se i nomi di file sono limitati a 14 caratteri, gzip.msdos.exe sarà compresso in gzi.msdx.exe.gz. I nomi non sono troncati nei sistemi che non hanno limite sulla lunghezza del nome del file.

Di default, gzip mantiene il nome e la time stamp (data e ora di creazione) originali del file nel file compresso. Queste sono usate quando si decompime il file con l'opzione -N. Ciò è utile quando il nome del file compresso è stato troncato o quando le date non sono state preservate dopo un trasferimento del file.

I file compressi possono essere riportati alla loro forma originale usando gzip -d o gunzip o zcat. Se il nome originale salvato nel file compresso non è adatto per questo file system, sarà costruito un nuovo nome da quello originale per renderlo legale.

gunzip accetta una lista di file nella riga di comando e rimpiazza ogni file il cui nome finisce con .gz, -gz, .z, -z, _z o .Z e che inizia con il magic number corretto con un file decompresso senza l'estensione originale. gunzip riconosce anche le estensioni speciali .tgz e .taz come abbreviazioni per .tar.gz e .tar.Z rispettivamente. Quando comprime, gzip usa se necessario l'estensione .tgz invece di troncare un file con estensione .tar.

gunzip attualmente può decomprimere file creati da gzip, ZIP, compress, compress -H o pack. Il rilevamento del formato d'ingresso è automatico. Quando si usano i primi due formati, gunzip verifica un CRC a 32 bit. Per pack, gunzip verifica la lunghezza del file decompresso. Il formato standard compress non è stato progettato per permettere verifiche sulla consistenza. Comunque gunzip è in grado qualche volta di rivelare un file .Z con errori. Se si ottiene un errore quando si decompime un file .Z, non si assuma che il file sia corretto solo perché l'uncompress standard non rileva l'errore. Ciò in genere significa solo che l'uncompress standard non verifica il suo input, e genera tranquillamente spazzatura in output. Il formato SCO compress -H (metodo di compressione lzh) non include un CRC ma permette comunque qualche verifica sulla consistenza.

I file creati da zip possono essere decompressi da gzip solo se contengono un unico file compresso con il metodo di "deflation". Questa possibilità è pensata solo per aiutare la conversione dei file tar.zip nel formato tar.gz. Per decomprimere file zip con più membri si usi unzip invece di gunzip.

zcat è identico a gunzip -c (su alcuni sistemi, zcat può essere installato come gzcat per preservare il link originale a compress). zcat decompime o una lista di file nella riga di comando o il suo standard input e scrive i dati decompressi nello standard output. zcat decomprimerà i file che hanno il magic number corretto che abbiano o meno il suffisso .gz.

Gzip usa l'algoritmo di Lempel-Ziv usato in zip e PKZIP. L'ammontare della compressione ottenuta dipende dalla dimensione dell'ingresso e dalla distribuzione delle sotto-stringhe comuni. Tipicamente, testi come codici sorgenti o inglesi sono ridotti del 60-70%. La compressione è generalmente molto migliore di quella ottenibile da LZW (usato in compress), codifica di Huffman (usata in pack), o codifica di Huffman adattativa (compact).

Sviluppatore	Jean-loup Gailly, Mark Adler
Ultima versione	1.12 (7 aprile 2022)
Tipo	<u>Compressione dei dati</u>
Compressione	<u>Lossless</u>
Estensione di	<u>Deflate</u>
Formato aperto?	sì
Sito web	<u>www.gzip.org</u> (<u>http://www.gzip.org</u>)

La compressione è sempre fatta, perfino se il file compresso è leggermente più grande del file originale. Il caso peggiore di espansione è di alcuni bit per il gzip file header, più 5 byte ogni blocco da 32K, o un rapporto di espansione del 0.015% per file più grandi. Si noti che il numero di disk block realmente usati non viene mai incrementato. gzip preserva il mode, le proprietà e le date dei file quando li comprime o li decomprime.

Formato del file

Gzip si basa sull'algoritmo Deflate che unisce LZ77 e la Codifica di Huffman. DEFLATE era previsto come un sostituto dell'algoritmo LZW e di altri algoritmi di compressione protetti da brevetti, che all'epoca limitavano l'usabilità di compress e di altri noti programmi di archiviazione.

Col termine *gzip* si fa spesso riferimento all'omonimo formato di file, che è composto da:

- una intestazione di 10 byte contenente un magic number, una versione del programma e un timestamp;
- intestazioni aggiuntive facoltative, come ad esempio il nome originale del file;
- un corpo centrale, contenente un insieme di dati compressi;
- 8 byte finali contenenti un checksum di tipo CRC-32 dei dati e la loro lunghezza originaria.

Sebbene questo formato di file permetta di concatenare più flussi di dati in ingresso (che sono decompressi e concatenati come fossero un unico flusso), *gzip* è solitamente usato per comprimere un unico file. Gli archivi compressi sono comunemente creati impacchettando una collezione di file tramite un programma di archiviazione come TAR e successivamente comprimendo l'archivio ottenuto. Il file `.tar.gz` o `.tgz` è comunemente detto un *tarball compresso*.

gzip non va confuso con il formato di file ZIP, il quale pure utilizza l'algoritmo DEFLATE. Il formato ZIP può contenere collezioni di file senza dover ricorrere ad un programma di archiviazione esterno, ma il risultato è meno compatto di quello ottenuto usando *gzip* in congiunzione ad un programma di archiviazione come ad esempio tar, in quanto i file sono compressi individualmente e quindi non ci si può avvantaggiare delle ridondanze esistenti tra file diversi (solid compression).

zlib è una libreria software che offre un'astrazione dell'algoritmo DEFLATE e che include nella sua API sia il supporto per il formato *gzip* che semplici funzionalità di gestione di flussi di dati compressi. Il formato di flusso zlib, DEFLATE e il formato di file *gzip* furono standardizzati rispettivamente nelle RFC 1950, 1951 e 1952.^{[8][9][10]}

Dato che l'intestazione del file in formato gzip contiene un timestamp, è necessario utilizzare programmi come zcmp o zdiff per determinare se due file compressi con *gzip* contengano in forma compressa gli stessi dati.

Altri usi

La versione 1.1 del protocollo HTTP prevede la possibilità di indicare un *content coding* per la compressione dati senza perdita. Gli header utilizzati sono Accept-Encoding e Content-Encoding, che possono assumere i valori "gzip" (o "x-gzip"), "compress" (o "x-compress") e "deflate".^{[11][12]}

Fin dai tardi anni '90, bzip2, una utility di compressione di file basata sull'algoritmo block-sorting, ha guadagnato una certa popolarità come sostituto del gzip. Produce file considerevolmente più piccoli (specialmente per codici sorgenti e altri testi strutturati), ma al costo di un maggiore consumo di memoria e

tempo di calcolo (Addirittura fino a 4 volte maggiore). I tarball bzip2-compressed sono solitamente chiamati `.tar.bz2`.

AdvanceCOMP ha una implementazione DEFLATE che produce file compatibili con gzip con una compressione migliore che lo gzip stesso.

Il corrispondente programma per scompattare file compressi con gzip è **gunzip**.

Note

- ¹ ^{([EN](#))} *GNU Gzip: General file (de)compression*, su *Progetto GNU*.
- ² ^{([EN](#))} Sahinalp e Rajpoot.
- ³ ^{([EN](#))} *The gzip home page*, su *gzip.org*.
- ⁴ ^{([EN](#))} *The GZIP File Format*, su *FileFormat.info*.
- ⁵ ^{([EN](#))} *GNU Gzip*, su *Progetto GNU*.
- ⁶ ^{([EN](#))} *gzip*, su *FreeBSD General Commands Manual*.
- ⁷ ^{([EN](#))} *gzip*, su *OpenBSD manual page server*.
- ⁸ ^{([EN](#))} *RFC 1950 — ZLIB Compressed Data Format Specification version 3.3*, su *Internet Engineering Task Force*.
- ⁹ ^{([EN](#))} *RFC 1951 — DEFLATE Compressed Data Format Specification version 1.3*, su *Internet Engineering Task Force*.
- ¹⁰ ^{([EN](#))} *RFC 1952 — GZIP file format specification version 4.3*, su *Internet Engineering Task Force*.
- ¹¹ ^{([EN](#))} *RFC 2616 — Hypertext Transfer Protocol -- HTTP/1.1*, su *Internet Engineering Task Force*.
- ¹² ^{([EN](#))} *RFC 7231 — Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*, su *Internet Engineering Task Force*.

Bibliografia


- ^{([EN](#))} S. Cenk Sahinalp e Nasir M. Rajpoot, *Dictionary-Based Data Compression: An Algorithmic Perspective*, in Khalid Sayood (a cura di), *Lossless Compression Handbook*, Academic Press, 2003.

Voci correlate

- Compressione dei dati
- Deflate
- tar (software)

- [zlib](#)

Altri progetti

-  [Wikimedia Commons \(https://commons.wikimedia.org/wiki/?uselang=it\)](https://commons.wikimedia.org/wiki/?uselang=it) contiene immagini o altri file su **[gzip](https://commons.wikimedia.org/wiki/Category:Gzip?uselang=it)** (<https://commons.wikimedia.org/wiki/Category:Gzip?uselang=it>)

Collegamenti esterni

-
- (EN) [Sito ufficiale](#), su [gnu.org](https://www.gnu.org).
- [Gzip / Gzip \(altra versione\)](#), su packages.debian.org.
- [Repository sorgenti di Gzip](#), su git.savannah.gnu.org.
- [Repository sorgenti di Gzip](#), su svnweb.freebsd.org.
- [Repository sorgenti di Gzip](#), su cvsweb.netbsd.org.
- (EN) [Gzip](#), su [Free Software Directory](https://www.free-software-directory.org).
- (EN) [RFC 1952 — GZIP file format specification version 4.3](#), su [Internet Engineering Task Force](#).
- (EN) Aaron M. Renn, [The gzip Recovery Toolkit](#), su urbanophile.com.



Portale Software libero: accedi alle voci di Wikipedia che trattano di software libero

Estratto da "<https://it.wikipedia.org/w/index.php?title=Gzip&oldid=134931751>"

Questa pagina è stata modificata per l'ultima volta il 12 ago 2023 alle 19:24.

Il testo è disponibile secondo la licenza Creative Commons Attribuzione-Condividi allo stesso modo; possono applicarsi condizioni ulteriori. Vedi le condizioni d'uso per i dettagli.