

Բ.Գ. ԱԹԱՅԱՆ

ՏՎՅԱԼՆԵՐԻ ԱՆԿՈՐՈՒՄՍ ՍԵՂՄԱՆ DEFLATE ԱԼԳՈՐԻԹՄԻ ՎՐԱ ՀԻՄՆՎԱԾ ԱՄՊԱՅԻՆ ՊԱՀՈՒՍԱՎՈՐՄԱՆ ՄԵԹՈԴԻ ՄՇԱԿՈՒՄԸ

Մշակվել է ամպային միջավայրում տվյալների պահուստավորման արդյունավետ մեթոդ՝ հիմնված անկորուստ սեղմման DEFLATE ալգորիթի վրա: Առաջարկվում է ձևաչափ, որը թույլ է տալիս կատարել մեծ տվյալների արագագործ պահուստավորում, ինչպես նաև ստանալ պահուստի փոքր ծավալ:

Առանցքային բառեր. ամպային պահուստավորում, անկորուստ սեղմում, արագագործ պահուստավորում:

Գիտատեխնիկական առաջընթացի և տեղեկատվական տեխնոլոգիաների զարգացման հետ մեկտեղ էքսպոնենցիալ կերպով աճում է քոմպյուտերային համակարգերում պահպանվող տվյալների ծավալը: Հաշվի առնելով այն փաստը, որ թվային տեղեկատվությունը մեր օրերում բավականին թանկարժեք է, արդիական խնդիր է դառնում այդ տվյալների պահուստային օրինակների ստեղծումը և պահպանումը: Ամպային տեխնոլոգիաների լայնածավալ տարածումը թույլ է տալիս նորովի լուծել պահուստավորված տվյալների պահպանման խնդիրը: Եթե նախկինում տվյալների պահպանման համար օգտագործվում էին միայն ֆիզիկական լոկալ կրիչներ (CD/DVD, USB memory, HDD memory և այլն), ապա ներկայումս յուրաքանչյուր ոք հնարավորություն ունի իր տվյալները պահպանել ամպում: Տրամաբանական է, որ ամպային տեխնոլոգիաների տրամադրած հնարավորությունները կարելի է արդյունավետ օգտագործել տվյալների պահուստավորման համար:

Պահուստավորում իրագործելիս նպատակահարմար է խմբավորել ֆայլերը (օրինակ՝ TAR ձևաչափով [1]), քանի որ ֆայլերի առանձին պահուստավորումը կապված է հավելյալ ժամանակային և հաշվողական ծախսերի հետ: Փորձը ցույց է տալիս, որ մեկ ֆայլի անվտանգությունը և ամբողջականությունը ապահովելն ավելի դյուրին է և, հետևաբար, քիչ ծախսատար:

Մեր օրերում տվյալների պահպանումը դժվար է պատկերացնել առանց սեղմման: Քանի որ պահուստավորվող տվյալների ծավալը կարող է լինել բավականին մեծ, ապա սահմանափակ ռեսուրսների պայմաններում ակնհայտ է, որ պահուստավորվող տվյալներն անհրաժեշտ է պահպանել սեղմված կերպով: Դա կարելի է իրագործել օգտագործելով որոշակի ֆայլային ձևաչափ, որը թույլ կտա կատարել ինչպես տվյալների խմբավորում, այնպես էլ տվյալների սեղմում:

Ներկայումս առավել հայտնի և շատ օգտագործվող սեղմում ապահովող ձևաչափերից են, օրինակ՝ ZIP, RAR, 7ZIP, GZIP և այլն: Հետազոտվել է մասնավոր-

րապես ZIP ձևաչափը [2] իր կիրառություններով: Այդ ընտրությունը պայմանավորված է լայնածավալ օգտագործվող ծրագրավորման լեզուներով գրված ZIP ձևաչափով աշխատելու միջոցներ տրամադրող գրադարանների (օրինակ՝ zlib, System.IO.Compression և այլն [3,4]) հասանելիությամբ: Ինչպես հայտնի է, սեղմման համար գոյություն ունեցող ալգորիթմները բաժանվում են երկու հիմնական տեսակի՝ կորուստներով և անկորուստ սեղմում: ZIP ձևաչափի հիմքում ընկած է անկորուստ սեղմման DEFLATE [5] ալգորիթմը, ինչը ZIP ձևաչափի ընտրության երկրորդ պատճառն է, քանի որ պահուստավորման ընթացքում տվյալների ամբողջականությունը հանդիսանում է բարձր կարևորություն ունեցող նախապայման:

ZIP ձևաչափն ունի օգտագործման 2 տարբերակ՝ ZIP32 և ZIP64: Սկզբում օգտագործվող տարբերակը եղել է ZIP32 ձևաչափը, որը որոշ ժամանակ անց սկսեց չբավարարել արդիական խնդիրների լուծմանը, մասնավորապես, ZIP32 ձևաչափի գլխամասային գրառումների ֆիքսված չափերի պատճառով մեկ պահուստում պահպանվող ֆայլերի քանակը խիստ սահմանափակ էր: Այնուհետև որոշվեց ZIP32 ձևաչափը փոխարինել ZIP64-ով: Քանի որ ZIP32-ը հայտնի էր և լայնորեն օգտագործվող, նոր ZIP64-ը պետք է ապահովեր հետադարձ համատեղելիություն իրեն նախորդող ձևաչափի հետ: Այդ պատճառով ZIP64 ձևաչափի գլխամասը ընդգրկեց ZIP32 ձևաչափում ներառված գրառումները, և այդ կրկնօրինակման արդյունքում գլխամասի չափը մեծացավ:

ZIP ձևաչափի հետազոտության արդյունքում ձևավորվեցին որոշակի խնդիրներ՝

- մեկ պահուստում պահվող ֆայլերի քանակի սահմանափակում,
- ձևաչափի գլխամասի չափի մեծացում:

Ներկայացված խնդիրներն առավել արդիական են մեծ տվյալների պահուստավորման պայմաններում, երբ պահուստավորվող տվյալների ծավալը կարող է հասնել տերաբայթերի, պետաբայթերի, իսկ քանակը՝ միլիոնների և միլիարդների: Եվ քանի որ ZIP ձևաչափը նպատակահարմար չէ մեր առջև դրված խնդիրները լուծելու համար, մշակվեց նոր ձևաչափ, որն առավել հարմարեցված է մեծ ծավալի տվյալների պահուստավորման գործընթացում օգտագործելու համար:

Պահուստի ֆայլի հիմնական կառուցվածքը հետևյալն է (նկ.)

- Files - պահուստավորված ֆայլերի ամբողջությունը՝ սեղմված վիճակում,
- Footer-վերջնամաս, նախատեսված CDR-ի (Central Directory Record) մասին, ինչպես նաև որոշ այլ տեղեկություններ պահելու համար,
- CDR, որը պարունակում է տեղեկություններ պահուստավորված ֆայլերի մասին:

Files			Footer				CDR							Offset to footer
Extra size	Extra	Compressed file	Version	Extra size	Extra	CDR size	CRC	Filename length	Filename	Offset to file	Compressed file length	Uncompressed file length	...	
4 byte	...	...	1 byte	4 byte		4 byte	4 byte	2 byte	...	8 byte	8 byte	8 byte	...	4 byte

Նկ. Մշակված ձևաչափի կառուցվածքը

Footer գլխամասի գրառումը իր հերթին պարունակում է հետևյալ դաշտերը՝

- Version - ձևաչափի ընթացիկ տարբերակի համարը,
- Extra size - ռեզերվային դաշտի գրառման ընդհանուր չափը,
- Extra - ռեզերվային դաշտ,
- CDR size - CDR դաշտի գրառման ընդհանուր չափը:

CDR գրառումը պարունակում է հետևյալ դաշտերը՝

- CRC - Cyclic Redundancy Check, ֆայլի ամբողջականության ստուգման համար նախատեսված հսկիչ արժեքը,
- Filename - պահուստավորված ֆայլերի անվանումները,
- Filename length - պահուստավորված ֆայլերի անվանումների երկարությունը,
- Offset to the start of file - պահուստավորված ֆայլի սկզբի վրա ցուցիչը,
- Compressed file length - սեղմված ֆայլի երկարությունը,
- Uncompressed file length - օրիգինալ ֆայլի երկարությունը:

Առաջարկվող ձևաչափով պահուստավորված տվյալների սեղմման համար առաջարկվում է օգտագործել անկորուստ սեղմման DEFLATE ալգորիթը [5], որը հաջողությամբ օգտագործվում է նաև համեմատվող ZIP ձևաչափում: Եթե գրառման դաշտի չափը L բայթ է, ապա պահուստում պահվող ֆայլի առավելագույն չափը կարելի է հաշվել հետևյալ կերպ.

$$N = 2^{L*8},$$

որտեղ N -ը պահուստում պահվող ֆայլի առավելագույն չափն է: Առաջարկվող ձևաչափում այս չափը հավասար է 2^{64} : Իսկ պահուստում պահվող ֆայլերի քանակը, ինչպես երևում է գլխամասի CDR գրառումից, անսահմանափակ է, սակայն ֆայլերի քանակի մեծացման հետ աճում է CDR գրառման չափը, և հետևաբար՝ դրա մշակման համար ծախսվող ժամանակը: Այս հատկություններով մշակված նոր ձևաչափը առավել հարմար է մեծ ծավալի տվյալների պահպանման համար, քան ZIP ձևաչափը: Բացի այդ, մշակված ձևաչափը իր գլխամասում չի պահում ավելցուկային կամ կրկնօրինակվող տվյալներ, այդ պատճառով մեծ պահուստների դեպքում ավելի արագագործ է, քան ZIP ձևաչափը: Պահուստավորված տվյալների պաշտպանությամբ

յունը ապահովելու համար մշակված ձևաչափում նախատեսված են հատուկ ռեզերվային դաշտեր, որոնք օգտագործվելու են պահուստի անվտանգության ապահովման համար, մասնավորապես, գաղտնագրային ալգորիթմի և բանալիի մասին տեղեկություններ պահելու համար:

Այսպիսով, հետազոտվել են ամպային միջավայրում մեծ տվյալների պահուստավորման խնդիրը, ինչպես նաև տվյալների պահուստավորման ներկայումս արդիական մեթոդները և պահուստային հայելիի ձևաչափերը: Արդյունքում՝ ստեղծվել է ամպային միջավայրում տվյալների պահուստավորման արդյունավետ ձևաչափ՝ հիմնված անկորուստ սեղմման DEFLATE ալգորիթմի վրա, որը թույլ է տալիս կատարել մեծ տվյալների արդյունավետ և արագ պահուստավորում:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. https://www.gnu.org/software/tar/manual/html_node/Standard.html
2. <https://pkware.cachefly.net/webdocs/casestudies/APPNOTE.TXT>
3. <http://www.zlib.net>
4. [https://msdn.microsoft.com/en-us/library/system.io.compression\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/system.io.compression(v=vs.110).aspx)
5. <http://www.zlib.net/feldspar.html>

Б.Г. АТАЯН

МЕТОД ОБЛАЧНОГО РЕЗЕРВНОГО КОПИРОВАНИЯ НА ОСНОВЕ DEFLATE АЛГОРИТМА СЖАТИЯ ДАННЫХ БЕЗ ПОТЕРЬ

Разработан продуктивный формат резервного копирования данных в облачных системах. Формат основан на DEFLATE алгоритме сжатия данных без потерь. Предложенный формат предоставляет быстрое резервное копирование больших объемов данных, в то же время обеспечивая небольшие размеры резервного файла.

Ключевые слова: облачное резервное копирование, сжатие без потерь, быстрое резервное копирование.

B.G. ATAYAN

THE CLOUD BACKUP METHOD BASED ON LOSSLESS DEFLATE DATA COMPRESSION ALGORITHM

The lossless data compression DEFLATE algorithm-based efficient file format has been developed for backup in Cloud environments. The proposed format provides fast backup of large amounts of data, and at the same time ensures a small backup file size.

Keywords: cloud backup, file format, lossless compression, fast backup.