ZIP101 an archive walkthrough

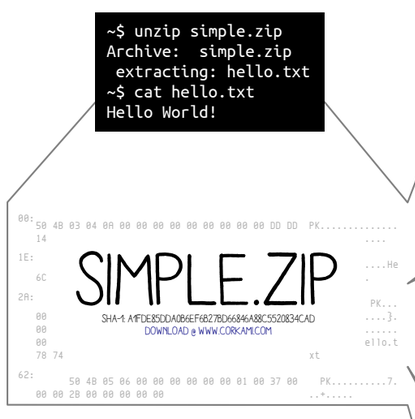
ANGE ALBERTINI
CORKAMI.COM

DISSECTED FILE

```
~$ unzip simple.zip
Archive:  simple.zip
 extracting: hello.txt
~$ cat hello.txt
Hello World!
```

SIMPLE.ZIP

SHA-1: A1FDE85DDA0B6EF6B27BD66846A88C5520834CAD
DOWNLOAD @ WWW.CORKAMI.COM



LOCAL FILE HEADER
ARCHIVED FILE INFORMATION

FILE DATA
ARCHIVED FILE CONTENT

CENTRAL DIRECTORY
LIST OF FILE HEADERS IN THE ARCHIVE

FILE NAME

END OF CENTRAL DIRECTORY

HEXADECIMAL DUMP	ASCII DUMP	DESCRIPTION*	VALUE
50 4B 03 04 0A 00 00 00 00 00 00 00 00 00 DD DD	PK.....	3 local file header signature	PK\x03\x04
14 7D 0D 00 00 00 0D 00 00 00 00 00 00 00	.}.....	version needed to extract	10 (default value)
		compression method	0 (no compression)
		crc-32	0x7d14dddd
		compressed size	0xd
		uncompressed size	0xd
Offset:0x1e 48 65	He	file data	"Hello World!\n"
6C 6C 6F 20 57 6F 72 6C 64 21 0A	llo.World!.		
Offset:0x2b 50 4B 01 02 00	PK...	2 central file header signature	PK\x01\x02
00 0A 00 00 00 00 00 00 00 00 00 DD DD 14 7D 0D	.....}.	version needed to extract	10 (default value)
00 00 00 0D 00 00 00 09 00 00 00 00 00 00 00 00	.....	crc-32	0x7d14dddd
00 00 00 00 00 00 00 00 00	.....	compressed size	0xd
		uncompressed size	0xd
		file name length	9
		relative offset of local header	0
Offset:0x59 68 65 6C 6C 6F 2E 74	hello.t	file name	hello.txt
78 74	xt		
Offset:0x62 50 4B 05 06 00 00 00 00 00 00 01 00 37 00	PK.....7.	1 end of central dir signature	PK\x05\x06
00 00 2B 00 00 00 00 00	..+.....	total number of entries in the central directory	1
		size of the central directory	0x37
		offset of start of central directory with respect to the starting disk number	0x2b

THIS IS THE WHOLE FILE, HOWEVER, MOST ZIP FILES CONTAIN MANY MORE ELEMENTS.
EXPLANATIONS ARE SIMPLIFIED, FOR CONCISENESS.

PARSING

THE ARCHIVE IS ANALYZED BACKWARD:

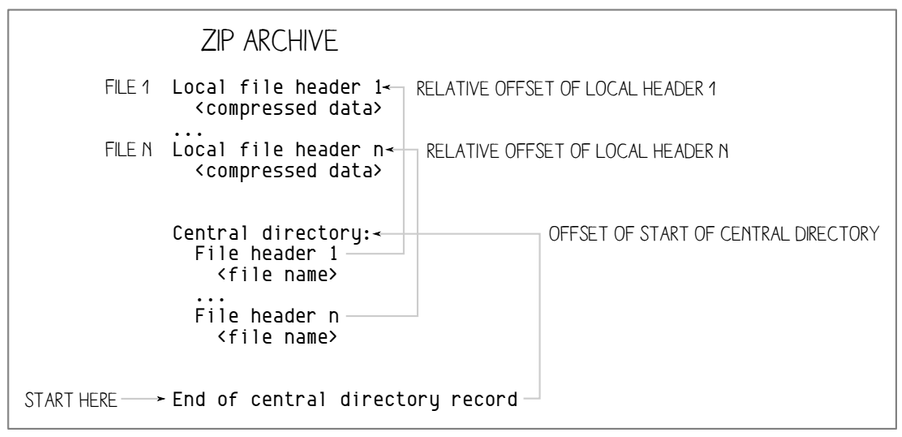
1 THE END OF CENTRAL DIRECTORY IS LOCATED (BY SCANNING), AND PARSED
2 THE CENTRAL DIRECTORY IS LOCATED, AND PARSED
3 EACH LOCAL FILE HEADER IS PARSED

ZIP ARCHIVE

```
FILE 1   Local file header 1   <- RELATIVE OFFSET OF LOCAL HEADER 1
           <compressed data>
         ...
FILE N   Local file header n   <- RELATIVE OFFSET OF LOCAL HEADER N
           <compressed data>

         Central directory:    <- OFFSET OF START OF CENTRAL DIRECTORY
           File header 1
             <file name>
           ...
           File header n
             <file name>

START HERE -> End of central directory record
```



TRIVIA

- THE ZIP FORMAT WAS CREATED BY PHILLIP WALTER KATZ, IN 1989
- *IN THE OFFICIAL SPECIFICATION, STRUCTURE NAMES ARE NOT FIXED, ONLY DESCRIBED.
- THE FILE LIST IS WRITTEN LAST TO MINIMIZE DISK SWAPS WHEN CREATING MULTI-VOLUME ARCHIVE ON THE FLY.
- THE FILES' SIZES AND CRCS ARE DUPLICATED TO BE ABLE TO LIST THE ARCHIVE WITHOUT REQUIRING ALL VOLUMES TO BE ACCESSED
- A ZIP FILE IS NOT ENFORCED TO START AT OFFSET 0.

VERSION 1.00
2013/12/24
CC BY