wiki:

<https://en.wikipedia.org/wiki/Zip_(file_format)>

<https://tools.ietf.org/html/rfc1951>

<https://users.cs.jmu.edu/buchhofp/forensics/formats/pkzip.html>

zip文件格式：



最后面是EOCD，用来记录有多少file entry和central directory开始的位置等。可以通过EOCD用来查找到central directory的位置。

EOCD前面的是central directory。每一个文件，在这里都有一个记录，里面有filename，文件数据开始的offset。可以通过central directory查找到文件关联数据的位置。

最开始的是文件定义，分为local file header和具体的data。local file header用来记录文件名字和文件压缩数据开始的地方。

central directory和local file header里面有重复的地方，如文件名和文件长度。一般已central directory里面的数据为准。但文件压缩数据开始的地方，还是要从local file header里面获取。

文件数据的存储方式一般有2种，**STORE**和**DEFLATE**。

STORE表述文件数据直接存储在文件里面，没有结果压缩。

DEFLATE表示文件数据经过压缩后再存储到文件里面，压缩算法一般是LZ77。

**EOCD field data:**

|  |  |  |
| --- | --- | --- |
| **Offset** | **Bytes** | **Description**[[25]](https://en.wikipedia.org/wiki/Zip_(file_format)#cite_note-appnote-25) |
| 0 | 4 | End of central directory signature = 0x06054b50 |
| 4 | 2 | Number of this disk |
| 6 | 2 | Disk where central directory starts |
| 8 | 2 | Number of central directory records on this disk |
| 10 | 2 | Total number of central directory records |
| 12 | 4 | Size of central directory (bytes) |
| 16 | 4 | Offset of start of central directory, relative to start of archive |
| 20 | 2 | Comment length (*n*) |
| 22 | *n* | Comment |

**Central Directory file header**

|  |  |  |
| --- | --- | --- |
| **Offset** | **Bytes** | **Description**[[25]](https://en.wikipedia.org/wiki/Zip_(file_format)#cite_note-appnote-25) |
| 0 | 4 | Central directory file header signature = 0x02014b50 |
| 4 | 2 | Version made by |
| 6 | 2 | Version needed to extract (minimum) |
| 8 | 2 | General purpose bit flag |
| 10 | 2 | Compression method |
| 12 | 2 | File last modification time |
| 14 | 2 | File last modification date |
| 16 | 4 | CRC-32 |
| 20 | 4 | Compressed size |
| 24 | 4 | Uncompressed size |
| 28 | 2 | File name length (*n*) |
| 30 | 2 | Extra field length (*m*) |
| 32 | 2 | File comment length (*k*) |
| 34 | 2 | Disk number where file starts |
| 36 | 2 | Internal file attributes |
| 38 | 4 | External file attributes |
| 42 | 4 | Relative offset of local file header. This is the number of bytes between the start of the first disk on which the file occurs, and the start of the local file header. This allows software reading the central directory to locate the position of the file inside the .ZIP file. |
| 46 | *n* | File name |
| 46+*n* | *m* | Extra field |
| 46+*n*+*m* | *k* | File comment |

**local file header**

|  |  |  |
| --- | --- | --- |
| **Offset** | **Bytes** | **Description**[[25]](https://en.wikipedia.org/wiki/Zip_(file_format)#cite_note-appnote-25) |
| 0 | 4 | Local file header signature = 0x04034b50 (read as a little-endian number) |
| 4 | 2 | Version needed to extract (minimum) |
| 6 | 2 | General purpose bit flag |
| 8 | 2 | Compression method |
| 10 | 2 | File last modification time |
| 12 | 2 | File last modification date |
| 14 | 4 | CRC-32 |
| 18 | 4 | Compressed size |
| 22 | 4 | Uncompressed size |
| 26 | 2 | File name length (*n*) |
| 28 | 2 | Extra field length (*m*) |
| 30 | *n* | File name |
| 30+*n* | *m* | Extra field |