# Huge documents in LibreOffice (Zip64 support)

Attila Szűcs
Software Engineer
attila.szucs@collabora.com


Collabora Online

Technical Day
COOL days

LibreOffice Technology

Technical Day
COOL days
LibreOffice Technology

# What is ZIP64 and why it is needed

- Most document formats are compressed with zip
- The original .ZIP file format was designed at year 1989
  old system – old limitations
  e.g. filesize stored in 32bit = max 4gb
  extensible
- technology advanced -> limitations reached
- Year 2001: ZIP64 extension

Note ZIP64 can be used even if it is not needed.

Collabora Online

Technical Day
COOL days
LibreOffice Technology

# ZIP64 new limitations

1) uncompressed file size: 32bit -> 64bit (4gigaB->16exaB)
2) compressed archive size: 32bit -> 64bit
3) file count: 16bit -> 32bit (64k->4g)
4) disk count: 16bit -> 32bit
5) Some less important internal limitations
   like size of the central directory

For most LibreOffice Documents, old ZIP limitations are enough, practically only uncompressed filesize can be problem for a while.

Collabora Online

# New/extended records to store data

- extra field
- Data descriptor

can be in "Local file header", and in "Central directory header"

Implemented export/import in LibreOffice

---

- zip64 end of central directory record
- zip64 end of central directory locator

Not Implemented in LibreOffice

ZIP64 mode is not just a flag for the entire archive.
Every record (or file) can be in (or not in) ZIP64 mode separately
Even 2 filesize data for the same file can be in different mode

Technical Day
COOL days
LibreOffice Technology

# The challenge

Zip allow a lot of special cases:

- files individually can be
  - compressed or not (multiple algorithms)
  - encrypted or not (multiple algorithms)
- even central directory can be encrypted
- streamed, split into segments, self extracting
- Lot of extensions

Zip standard not always exact, it allows many possibilities, but hard to prepare for every use case.

Technical Day
COOL days
LibreOffice Technology

# Test case

Unittest: small zip64 files that is fast to load.

For 4gb+ (content.xml) size → manual test only

It works, but slow. (what would we expect from a huge doc.)

Release: several minutes to import
Debug: it was like 40 minutes for me
Export nearly the same.

Collabora Online

# Future possibilities

Compressed size 32bit→64bit  (Partially implemented)

What need:

- load/save zip64 end of central directory record / locator
- Make sure all usage is 64bit compatible
  1 local sal_uInt32 variable, function parameter, or return value can break everything

Collabora Online

Technical Day COOL days

LibreOffice Technology

# Thank you!

*By Attila Szűcs*

@CollaboraOffice
hello@collaboraoffice.com
www.collaboraoffice.com