

Copy Activity: Unsupported compression causes file corrupted

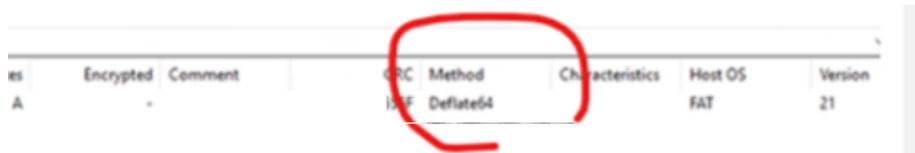
Last updated by | Veena Pachauri | Mar 8, 2023 at 11:22 PM PST

Contents

- [Issue](#)
- [RCA](#)
- [More detail](#)

Issue

Customer is trying to unzip a file that is stored in a blob container. I have a pipeline with a single Copy activity that has a source configured with compression set to deflate64(Or any unsupported type). This activity runs with success and produces the text file contained within the zip file HOWEVER there is a problem with the text within the file. It appears corrupted. When I unzip this file locally, it is fine.



| es | Encrypted | Comment | CRC | Method | Characteristics | Host OS | Version |
|----|-----------|---------|------|-----------|-----------------|---------|---------|
| A | - | | 13FF | Deflate64 | | FAT | 21 |

RCA

Customer's zip file was compressed by the algorithm of "deflate64", while our internal zip library only supports "deflate".

If the zip file is compressed by windows system, and the overall file size exceeds a certain number (maybe 2 GB I don't remember clearly), windows will use deflate64 by default, resulting in not supported by ADF. On the other hand, if the file size is smaller, or use some 3rd party zip tools which supports specifying the compress algorithm, it will be "deflate".

More detail

Actually, both public docs <https://docs.microsoft.com/en-us/azure/data-factory/format-binary> and <https://docs.microsoft.com/en-us/azure/data-factory/format-delimited-text> clearly say we do not support deflate64 format.

| | |
|------------------|--|
| compressionCodec | <p>The compression codec used to read/write text files.</p> <p>Allowed values are bzip2, gzip, deflate, ZipDeflate, TarGzip, Tar, snappy, or lz4. Default is not compressed.</p> <p>Note currently Copy activity doesn't support "snappy" & "lz4", and mapping data flow doesn't support "ZipDeflate", "TarGzip" and "Tar".</p> <p>Note when using copy activity to decompress ZipDeflate/TarGzip/Tar file(s) and write to file-based sink data store, by default files are extracted to the folder: <code><path specified in dataset>/<folder named as source compressed file>/</code>, use <code>preserveZipFileNameAsFolder/</code> <code>preserveCompressionFileNameAsFolder</code> on copy activity source to control whether to preserve the name of the compressed file(s) as folder structure.</p> |
|------------------|--|

How good have you found this content?

