Xilinx Gzip Compression User Guide

(Version 1.0)

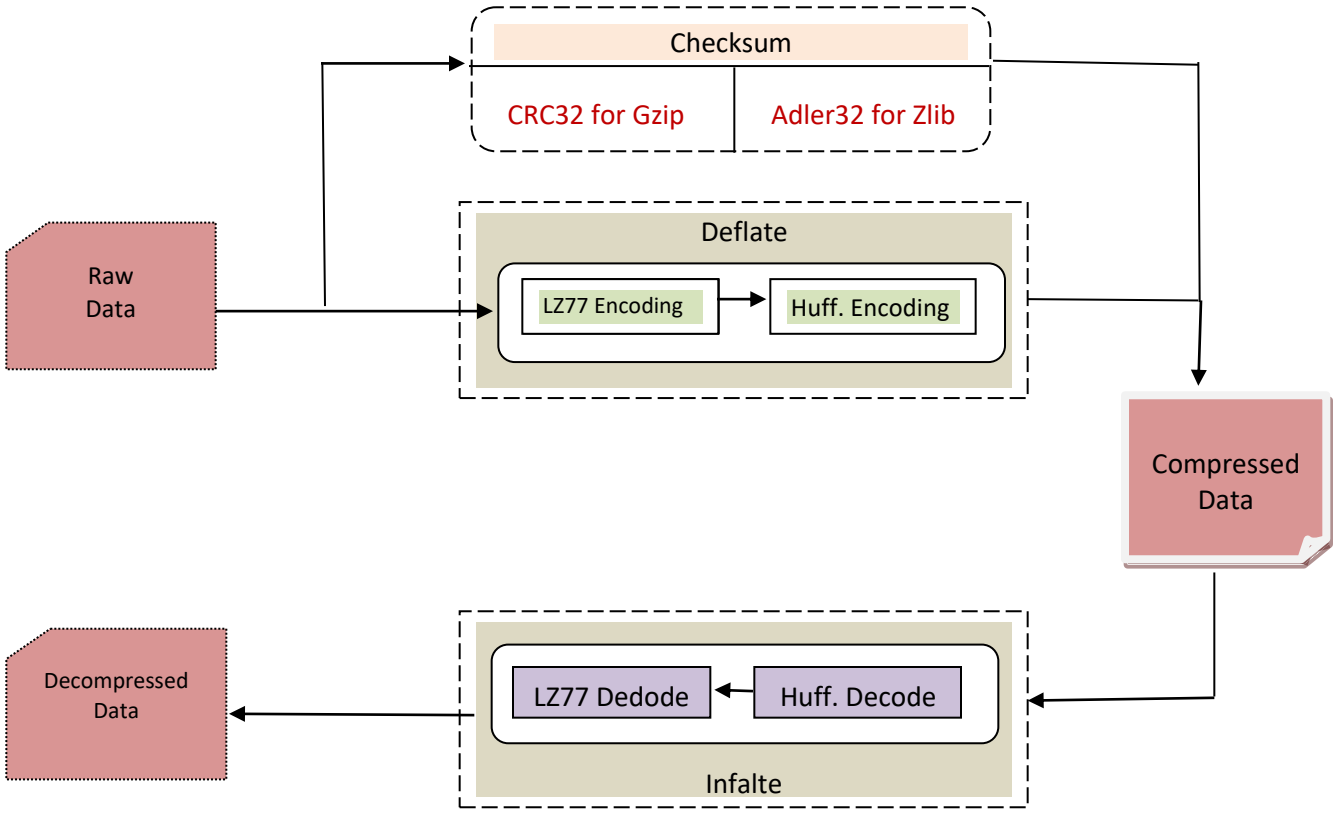
1. Introduction

Xilinx Gzip compression application is a lossless data compression, based on the standard gzip application which is based on the DEFLATE algorithm, which is a combination of LZ77 and Huffman coding. The application supports both compression and decompression with various other features, it also supports the zlib compression and decompression.

This application can run on machine with Alveo U50 (shell: xilinx_u50_gen3x16_xdma_201920_3).

1.1 Application brief architecture

- LZ77 compression algorithm works by using a sliding window to find sequences of data that are repeated, and encoding each repeated sequence by a pair of numbers called a length-distance pair.
- Huffman encoding is a statistical compression method. It encodes data with variable-length codes, and lengths of the codes are based on the frequencies of corresponding symbols.



2. Application Usage

The application is containerized and can be easily run in a few minutes in the Nimbix cloud or on premises.

2.1 Host Options

The application executable will be added as an environment variable as part of the docker image, to explore the various supported command line options, you can type: **xgzip -h**

```
================================================================
Usage: ./xgzip.exe [Options] [Files]

        --help,             -h        Print Help Options
        --compress,         -c        Compress
        --decompress,       -d        DeCompress
        --test,             -t        Compress Decompress
        --c_file_list,      -cfl      Compress list files
        --d_file_list,      -dfl      Decompress list files
        --file_list,        -l        List of Input Files
        --zlib,             -zlib     [0:GZIP, 1:ZLIB]          Default: [0]
        --ck,               -ck       Compress CU [0-7]         Default: [0]
        --dk,               -dk       Decompress CU [0-6]       Default: [0]
        --max_cr,           -mcr      Maximum CR                Default: [10]
        --verbose,          -v        Detailed output [0|1]     Default: [0]
```

Fig 1. Executable options

2.2 Sample Commands for Compression & Decompression

Download sample file with the command: **wget http://sun.aei.polsl.pl/~sdeor/corpus/nci.bz2**

- **Compress a file**

  Gzip compress: **xgzip -c <path_to_input_file>**

  Zlib compress: **xgzip -c <path_to_input_file> -zlib 1**

  The output file will be <file_name>.gz for gzip and <file_name>.xz for zlib

  Compress sample file**: xgzip -c nci.bz2**

- **Decompress a file**

  Xilinx Decompress**: xgzip -d <path_to_compressed_file>**

  Decompress sample file**: xgzip -d nci.bz2.gz**

- **Test a file with Xilinx compression and Decompression**

  Test the sample file**: xgzip -t nci.bz2**

  To see the detailed output run**: xgzip -t nci.bz2 -v 1**

3. Prerequisite

3.1 Device and Software

This application supports Xilinx FPGA Alveo U50 card at this moment. To run this application on users' machines, please make sure:

- Xilinx FPGA Alveo U50 (shell xilinx_u50_gen3x16_xdma_201920_3) card is installed correctly.
- Docker (with sudo access)

When deployed in Nimbix, PushToCompute flow will deploy the application in an instance with ubuntu18.04, U50, and XRT 2020.1.

4. Run Application

Below are the supported options to run the application.

Table 1. lists the entries user can use in this application.

Option Description	Options	Description	Default
--help	-h	Print help options	NA
--compress	-c	Compress a file	NA
--decompress	-d	Decompress a file	NA
--test	-t	Test a file with both compression and decompression	NA
--c_file_list	-cfl	Compress list of files	NA
--d_file_list	-dfl	Decompress list of files	NA
--file_list	-l	List of input files	NA
--ck	-ck	Compress compute unit [0-7]	0
--dk	-dk	Decompress Compute unit [0-6]	0
--zlib	-zlib	[0:GZIP, 1:ZLIB] :switch to run ZLIB, default is GZIP	0
--max_cr	-mcr	Maximum CR	0
--verbose	-v	Detailed output	0

Table 1. Command list

5. *Performance Spec*

The below overall throughput is calculated using the multiple file list on multiple core.

	Speed/Filesize	Total Cores	Overall speed
Compress	668.5 MB/s	8	5.3 GB/s
Decompress	678.4 MB/s	7	4.7 GB/s

Table 2. Performance table

Below are the file list run examples:

Reference for the silesia file list: http://sun.aei.polsl.pl/~sdeor/index.php?page=silesia

- Gzip list file run:

  Command used: xgzip -l silesia_fileset.list -v 1 (Refer table: 2.a and 2.b)

- Zlib list file run:

  Command used: xgzip -l silesia_fileset.list -v 1 -zlib 1 (Refer table: 3.a and 3.b)

Xilinx Gzip Compress			
E2E(MBps)	**CR**	**File Size(MB)**	**File Name**
496.348	2.551	9.971	sileisa/mr
480.805	4.962	5.345	sileisa/xml
463.760	1.810	6.152	sileisa/ooffice
498.588	2.860	6.627	sileisa/reymont
496.815	1.283	7.252	sileisa/sao
537.685	2.260	10.192	sileisa/dickens
600.745	2.438	51.220	sileisa/mozilla
625.651	7.482	33.553	sileisa/nci
543.117	1.976	10.086	sileisa/osdb
599.850	3.342	21.606	sileisa/samba
601.972	2.812	41.459	sileisa/webster
510.125	1.290	8.474	sileisa/x-ray

Table 2.a. Gzip compression list

Xilinx Gzip Decompress		
E2E(MBps)	**File Size(MB)**	**File Name**
200.487	3.909	sileisa/mr.xe2xd.gz
263.802	1.077	sileisa/xml.xe2xd.gz
184.567	3.399	sileisa/ooffice.xe2xd.gz
259.463	2.317	sileisa/reymont.xe2xd.gz
174.166	5.651	sileisa/sao.xe2xd.gz
260.031	4.510	sileisa/dickens.xe2xd.gz
295.197	21.005	sileisa/mozilla.xe2xd.gz
572.375	4.485	sileisa/nci.xe2xd.gz
218.906	5.103	sileisa/osdb.xe2xd.gz
366.631	6.464	sileisa/samba.xe2xd.gz
386.328	14.746	sileisa/webster.xe2xd.gz
159.906	6.570	sileisa/x-ray.xe2xd.gz

Table 2.b. Gzip Decompression list

Xilinx Zlib Compress			
E2E(MBps)	**CR**	**File Size(MB)**	**File Name**
521.691	2.551	9.971	sileisa/mr
480.351	4.962	5.345	sileisa/xml
459.104	1.810	6.152	sileisa/ooffice
499.445	2.860	6.627	sileisa/reymont
478.162	1.283	7.252	sileisa/sao
535.448	2.260	10.192	sileisa/dickens
594.933	2.438	51.220	sileisa/mozilla
626.481	7.482	33.553	sileisa/nci
539.138	1.976	10.086	sileisa/osdb
590.579	3.342	21.606	sileisa/samba
603.491	2.812	41.459	sileisa/webster
521.834	1.290	8.474	sileisa/x-ray

Table 3.a. Zlib Compression list

Xilinx Zlib Decompress		
E2E(MBps)	**File Size(MB)**	**File Name**
203.472	3.909	sileisa/mr.xe2xd.xz
263.152	1.077	sileisa/xml.xe2xd.xz
184.708	3.399	sileisa/ooffice.xe2xd.xz
259.594	2.317	sileisa/reymont.xe2xd.xz
169.811	5.651	sileisa/sao.xe2xd.xz
258.727	4.510	sileisa/dickens.xe2xd.xz
295.769	21.005	sileisa/mozilla.xe2xd.xz
583.182	4.485	sileisa/nci.xe2xd.xz
223.949	5.103	sileisa/osdb.xe2xd.xz
372.919	6.464	sileisa/samba.xe2xd.xz
382.145	14.746	sileisa/webster.xe2xd.xz
156.864	6.570	sileisa/x-ray.xe2xd.xz

Table 3.b. Zlib Decompress list