High Performance Computing
LP1 - Mini Project

Problem Statement :
Generic Compression : Run length encoding concurrently on many core GPUs.

Team Members :
41250 - Sahil Patil
41254 - Pralhad Kulkarni
41258 - Srusti Raybhoge

Abstract :
Compression allows scalable storage of large amounts of data and alleviates the I/O bottleneck for data-intensive applications. Over the years, a large number of compression algorithms have been developed to support various data types. Our goal is to develop a generic tool that helps quickly and accurately decide what compression scheme that best matches the constraints and expectations of users with respect to factors such as compression time(encoding/decoding), space (data/meta-data), and quality(lossiness).

Hardware and Software requirements :
CUDA, Google colab, GPU, c++ compiler

Introduction :
Generic compression, source coding or bit-rate reduction is the process of encoding information using fewer bits than the original representation.The process of reducing the size of a data file is often referred to as data compression. Compression is useful because it reduces resources required to store and transmit data. Data compression is critical to making large datasets scalable.

Parallel computing is a type of computing architecture in which several processors simultaneously execute multiple, smaller calculations broken down from an overall larger, complex problem.The primary goal of parallel computing is to increase available computation power for faster application processing and problem solving.

A graphics processing unit (GPU) is a specialized, electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. GPUs are used in embedded systems, mobile phones, personal computers, workstations, and game consoles. Modern GPUs are very efficient at manipulating computer graphics and image processing.

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Scope :
High performance computing is a form of digital computation that allows users to tackle and solve complex problems at a faster rate than standard computation. High performance computing focuses on efficiency, making more with less.

Objective :
Develop a generic tool that helps quickly and accurately decide what compression scheme(s) best match the constraints and expectations of users with respect to factors such as compression time(encoding/decoding), space (data/meta-data), and quality(lossiness).

Test Cases :

Input	Output	Remark
Original Length = 32535	**Compressed Length = 2843**	**Successfully compressed**

Results/ Output screenshots :

Original Length : 32628

Compressed Length : 2843
Compression achieved : 12.43545

Fig 1. Output

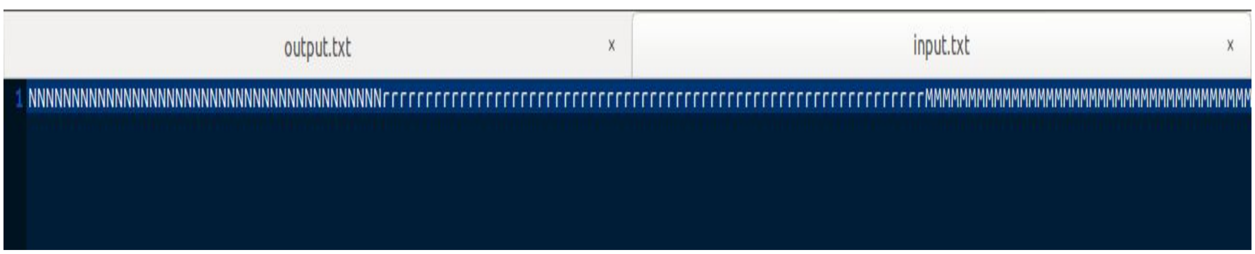

Fig 2. Input text file

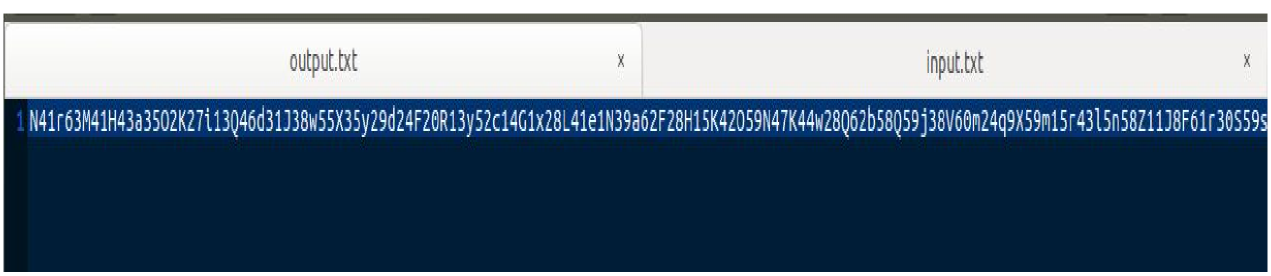

Fig 3. Output Text file

Conclusion :

Compression algorithm performance varies dramatically based on input data and parameter selection. We were successfully able to compress a large text file.