**Compression**

**What is data compression?**

The reduction in the size or number of bits of data needed to accurately represent the original data

* Commonly referred to as compression algorithm
  + Zip
  + Rar

Why is this needed?

* Save storage
* Speed up data transfers
* Decrease cost for hardware as well as the network speed itself

Important notes

* Don’t compress something that’s already been compressed
  + You can either lose important data OR nothing changes – pointless either way
* Important to reduce space from common things like log files
  + Contain a lot of redundant information so compression is necessary

**Two types**

* Lossy
* Lossless
  + LZ77
  + LZ78

**Lossless vs Lossy**

**Lossy**

A compression that may cause some data to be lost

* Lost or changed

Common in graphics, videos and images where the loss of a bit or will not drastically the file or how it works

For other files (**drawbacks**)

* Could cause an app not to run
* Or when compressing logs, there could be bits of data missing

**Lossless**

Compression if data that will never be lost and can be converted back to its original state

* Perfect for executables, text, spreadsheets etc.
* No room for error in these files
* Data loss in these types of files can drastically change functionality

Common lossless algorithms: LZ77 & LZ78

**LZ77**

Replaces repeated occurrences of data with a reference to a single master copy

* Normally held at beginning of file
* Consistently holds a set amount of data in memory that it last saw to compare matches

Also known as a Sliding Window Compression

* Amount of data held in memory is called The Sliding Window

Uncompressing this will place a copy of the master data in place of the references until all the references are replaced with the original data

**LZ78**

Similarly looks for repeated data

But replaces all repeated data with a reference to an internal dictionary

* Dictionary[….] = {index\_number, character}

Uncompressing this will replace the character at the correct index instead

* Will give exact output as the input