# Reflective writing in the theory/practice transfer paper

Matriculation number: 8240

Accepted topic of the transfer paper: Effective cold-storage of CI pipeline artifacts

These questions are intended for reflection and support during the preparation of your transfer paper. We recommend you to ask yourself recurrently these questions before, during, and after the elaboration of the transfer paper in order to develop a scientifically founded paper. Please answer the questions with one or two sentences (enumerations are also possible) and do not forget to refer to the corresponding chapter in your transfer paper. In contrast to the transfer paper and to the assignment clarification, writing in the first person is desired within the framework of reflective writing. This form is an integral part of the assessment of the theory/practice transfer paper and therefore must be included for evaluation along with the paper itself.

You can use 2000 characters (without spaces) as a guideline for answering the questions.

## Which are the central research questions of my topic?

The central research questions revolve around the topic of how to effectively compress CI pipeline artefacts at rest but retain efficient access.

This topic is split into two more detailed research questions in the introduction.

## Which objective do I want to achieve with the elaboration of the transfer paper and to what extent did I accomplish this at the end?

The goal is to find a solution to the at-rest compression challenges outlined and find the best-suited algorithm. Solving these allows a more efficient use of the available disk space and allows partial offloading of the storage requirements to the CPU (practically converting disk space and performance requirements into CPU performance requirements).

While compression algorithms always pose a trade-off between CPU requirements and compression ratio, a comprehensive overview has been provided. Based on this it is possible to make informed decisions on what algorithm to employ in a real-world system.

Additionally, the random-access challenge has been solved. While the solution presents yet another trade-off, a solid recommendation has been made regarding the parameters.

## How are my central questions/topics and the content that I have learnt during my academic studies connected?

The primary connection stems from the structural analysis of the underlying principles that serve most modern compression algorithms. Using this understanding, it was possible to derive potential methods to allow random access to algorithms.

Secondary connections are present in the statistical methods used.

## What information have I drawn on for my work and why have I chosen this source?

This research paper mostly relies on whitepapers and RFCs for the respective algorithms covered. For some algorithms, no such documentation was available and other sources were consulted (in the special case of Bzip this was a reverse engineered specification).

In addition to papers, many of the results obtained rely on a empirical of the algorithms in question.

## What is my approach to answering the research question?

The first research question is answered through empirical methods. In addition, the original papers for each algorithm (where available) have been consulted to gain insights into the methods used by the algorithms.

The second research question has been answered through the use of two methods. At first, we argumentatively outlined possible methods to allow random-access based on the knowledge acquired from research papers. To optimise these methods (namely the block size), another empirical analysis has been performed.

## Where do you see my personal contribution to answering the research question?

While the general topic of compression algorithm performance has been covered frequently in a general sense, few papers are concerning themselves with the specialised field of CI pipeline artefact compression. In this regard, the analysis provides a more tailored insight. Additionally, it covers the topic of random access. This is highly important when applying at-rest compression to CI artefact storage.

Additionally, this paper provides a comprehensive overview of modern compression algorithms that have been released within the last few years (namely Zstd, Brotli and Snappy).

## What have I learnt from writing this transfer paper and what would I do differently next time? A reference to a chapter is optional in this case.

The methodology did not account for the thermal state of the system running the analysis. In order to get more precise and repeatable results, the system should either be brought into a thermal steady state prior to testing or fixed to a certain frequency (effectively countering thermal throttling). However, this requires a longer test setup and more control over variables which were not accessible at the time (macOS does not allow you to deactivate thermal throttling and the thermal solution is not sufficient enough to maintain a constant frequency under heavy load — a more specialised hardware system would be required).

Upload the completed reflective writing form (as a separate document) together with your transfer paper in the Campus Information System (CIS).