LAB #2

In this lab, you will review the data collected from Lab #1, identify what information is relevant and useful, and write a script to preprocess the data. The goal is to clean and prepare the dataset.

01 Write a script that truncates your data to a maximum of MAX_TOKENS = 40 000.

02 Remove the following fields:

- global_seq_num, seq_num, thread_id, mach_time, schema_version, version, action, action_type
- All nested stat fields like st_size, st_gid, st_uid, st_mtimespec etc.
- parent_audit_token, responsible_audit_token, group_id, session_id
- tty, codesigning_flags, start_time