

Potential improvements to speed to queries:

Statement 1: Currently loading CSV without headers so that a second query must be used to delete the node containing the header information.

Improvement 1: Find out a way to load a CSV with header independent of header tags so that it is not needed to query a second time just to delete the node containing the header information.

Fixed 1 (Slightly): Only deleted from users, no need to have more queries for the other entity nodes.

Statement 2: Currently using match statements where exact information is provided. This is the closest way that has been approached where it will not take too much time.

Improvement 2: Include start/index/constraints to make queries much faster (unknown how to implement this thus far). After researching online, Batch import is a much more efficient way to load rows from a CSV in quantities of large datasets in little to no time compared to LOAD CSV.

Fixed 2 (Slightly): Created indexes on entity nodes to make it faster to match/merge – unknown whether or not this makes an impact.

Statement 3: Currently committing periodically every 10,000 rows to offset memory usage.

Improvement 3: Can commit periodically at a higher incremental state, however this is dependent on how much memory, and CPU performance the computer possesses to run efficiently.

Statement 4: Currently using actors to run the cypher commands for creating the skills, interests, projects, and organization entities.

Improvement 4: Cannot make actors more efficient to run the query without improving memory capacity, and/or CPU performance.