

Yulun Gao

Yizhou Pan

CS122B Project3

Performance Tuning report

For Parsing and Inserting batch of a huge amount of data we performed the following optimizations:

Batch inserts: Rather than submitting and committing one record at a time, such as insert only one movie with its information at one time, we performed batch inserts. Batch inserts method saves us preprocessing time on the database. This optimization was implemented from the start; thus, it was not possible to get the actual time improvements over not using batch inserts method. Furthermore, we were able to reduce the overhead that the database introduces when writing to the disk with each insert with auto commit turned off.

Caching: The new values inserted to the database, such as movies, stars and genres, all contained values, such as movies and stars generated their unique ID's, that were needed in further execution of the code. Rather than querying the database for these amount of values, we actually returned them upon insertion and were able to keep them at local for later reference.