

Abstrak

Extraction, Transformation, dan Loading, yang lebih dikenal dengan singkatan ETL, merupakan salah satu proses yang sangat penting dalam penerapan konsep *data warehouse*. Secara garis besar tugas dari proses ETL adalah untuk mengumpulkan data yang berasal dari sumber data dengan sistem operasi dan *environment* berbeda-beda menjadi sebuah kumpulan besar data yang terintegrasi. Dengan demikian semakin besarnya jumlah data yang akan diintegrasikan maka akan semakin berat dan semakin lama pula kinerja dari proses ETL. Hal tersebut akan mempengaruhi performansi *data warehouse* keseluruhan. Dalam buku tugas akhir ini penulis melakukan optimasi terhadap ETL dengan representasi *state-space* dan algoritma *heuristic-greedy* yang dapat mengurangi jumlah baris data yang diproses dalam setiap transformasi. Kemudian dilakukan analisis pada hasil optimasi ETL dengan membandingkan antara *cost value* dan waktu eksekusi yang dihasilkan algoritma *heuristic-greedy* dan algoritma *exhaustive*. Hasil akhir dari penelitian tugas akhir ini menunjukkan waktu eksekusi algoritma *heuristic-greedy* jauh lebih sedikit dibandingkan dengan algoritma *exhaustive*, dan algoritma *heuristic-greedy* menghasilkan nilai *cost* minimal yang sedikit lebih besar atau sama dengan nilai *cost* minimal yang dihasilkan algoritma *exhaustive*. Maka dari itu algoritma *heuristic-greedy* memberikan solusi yang sangat baik, karena dapat menghasilkan nilai *cost* minimal yang cukup memuaskan jika dibandingkan dengan hasil dari algoritma *exhaustive* dengan waktu yang jauh lebih cepat.

Kata kunci : etl, optimasi etl, *data warehouse*, *heuristic-greedy*, *state-space*, *cost model*