

## **Abstract**

Extraction, Transformation, and Loading, also known as ETL, is an important process in implementation of data warehouse concept. In general, ETL's task is to collect data from various sources with different environments and operation systems to become an integrated data collection. Thus, the greater size of data, the greater workload and execution time of ETL process, and that will influence all over datawarehouse performance. In this book, the author performs an ETL optimization with state-space representation and heuristic-greedy algorithm that can reduce the number of rows of data that are processed in each transformation. Afterward, the results of optimization are analyzed by comparing the cost value and execution time between heuristic-greedy and exhaustive algorithms. The end result of this research shows the execution time of heuristic-greedy algorithm is much less than the exhaustive algorithm, and heuristic-greedy algorithm generates minimal cost value slightly greater than or equal to the value of minimal cost that generated by exhaustive algorithm. Thus heuristic-greedy algorithm provides an excellent solution, because it can generates a minimum cost value that quite satisfactory when compared with the results of the exhaustive algorithm with much faster time.

keywords : etl, etl optimization, data warehouse, heuristic-greedy, state-space, cost model