Abstract

Data warehouse are present and developed as a data repository that is currently required by the company. amount of data thatmust be managed and spread as well as information that can be extracted from historical data into the purpose of the data warehouse creation. As a repository of data, data warehouse would have always interacted with a large amount of data, handle the transaction data in large numbers and high intensity. Therefore, the data warehouse is claimed to be able to handle every transaction that occurs in the operational database and capture data changes from each transaction and save it as a historical data with minimal delay. To deal with these problems, there is a process in the data warehouse that is called change data captue (CDC). CDC process will identify and process any changes that occur in the operation of the database are then transmitted and stored in temporary storage, called the staging database. With the number of data and transactions that occur each day, data warehouse needed a good performance from the CDC. To get a good performance, CDC certainly needed a good performance from each component of the building process of the CDC. In this final project, will performed the implementation and testing of the CDC to see the resulting performance based on CDC's execution time. Then later in the analysis of any components that play a role in the CDC and how the performance of any of the component. With the analysis of performance and find out the factors that influence performance, then the process of waiting in CDC can be eliminated or reduced the intensity and time. So the CDC process can generate a delay as small as possible and become a real time data warehouse.

Keywords: *data warehouse, change data capture, source database, staging database, asynchronous distributed hotlog*.