Abstract

Data Cleansing is an activity involving a process of detecting and correcting the errors and inconsistencies in data warehouse. It deals with identification of corrupt and duplicate data inherent in the data sets of a data warehouse to enhance the quality of data. The study looked into investigating some research works conducted in the area of data cleansing. A thorough review into these existing works was studied to determine the achievable goals and the limitations that arose based on the approaches conducted by the researchers. They identification of errors by most of these researchers has led into the development of several frameworks and systems to be implemented in the area of data warehousing. Generally, these findings will contribute to the emerging empirical evidence of the strategic role data cleansing play in the growth of organizations, institutions and other government agencies in terms of data quality and reporting purposes and also to gain competitive advantage since they will overcome the mere existence of dirty data.

References

- Heiko Muller, Johann-Christoph Freytag. (2003). Problems, Methods, and Challenges in
Comprehensive Data Cleansing, pp. 21.


Index Terms

Computer Science
Information Sciences

Keywords
Data inconsistency identification of errors organization growth