Abstract

Data errors occur in various ways when data is transferred from one point to the other. These data errors occur not necessarily from the formation/insertion of data but are developed and transformed when transferred from one process to another along the information chain within the data warehouse infrastructure. The main focus for this study is to conceptualize the data cleansing process from data acquisition to data maintenance. Data Cleansing is an activity involving a process of detecting and correcting the errors and inconsistencies in data warehouse. Poor data or "dirty data" requires cleansing before it can be useful to organizations. Data cleansing therefore deals with identification of corrupt and duplicate data inherent in the data sets of a data warehouse to enhance the quality of data. The research was directed at investigating some existing approaches and frameworks to data cleansing. The research attempted to solve the gaps identified in some data cleansing approaches and came up with a conceptual framework to overcome the weaknesses which were identified in those frameworks and approaches. This novel conceptual framework considered the data cleansing process from the point of data is obtained to the point of maintaining the data using a periodic automatic cleansing approach.
References

- Dongre Kuldeep (2004). Data cleansing strategies, pp. 10

Index Terms

Computer Science

Information Sciences

Keywords

Conceptual Framework    data cleansing process    gap analysis    dirty data