Abstract

In Data Warehouse (DW) environment, Extraction-Transformation-Loading (ETL) processes consume up to 70% of resources. Data quality tools aim at detecting and correcting data problems that affect the accuracy and efficiency of data analysis applications. Source data imported into the data warehouse often has different quality, format, coding etc. In order to bring all the data together in a standard, homogeneous environment, Extraction–transformation–loading (ETL) tools are used. ETL solutions provided so far are either proprietary and have limited functionality. Small and Medium Scale Enterprises (SME) and Small Scale Enterprises (SSE) cannot afford the licensing cost of these paid tools. The developed tool is capable of providing an integrated and open source data quality solution - MaSSEETL is to deal with naming conflicts, structural conflicts, date conversions, missing values and changing dimensions. MaSSEETL solves the appropriate errors with appropriate level of warning. In this paper, we are presenting the working of MaSSEETL. The tool provides an pragmatic evidence of strategic intensification of quality data in the academic and business enterprises.
An Open Source ETL Tool - Medium and Small Scale Enterprise ETL (MaSSEETL)


Index Terms

Computer Science

Information Sciences
Keywords
Data inconsistency  identification of errors  organization growth  ETL  data quality