

A review of the literature indicates that there are hundreds of tools that can be classified as data extraction, loading and cleansing tools. Only a small percentage of the tools perform the data cleansing and auditing functions, the majority perform extraction and loading functions. The majority of the auditing and cleansing tools were evaluated for this paper, as well as a small sample of the extraction and loading tools. Phase I of the literature review generated a preliminary list of questions for the data. Phase II refined the list and used it as the foundation for building a matrix to be used in evaluating data sources.

The original plan was to obtain copies of the software tools for testing. This proved to be very difficult. Only one vendor had a demo version available. The other vendors had marketing demos, which showed screen shots of the software, but did not allow testing the software with live data. The opportunity to talk to the vendors, and see the software, became available at the DCI Data Warehouse Conference in New York City. The next step was to talk to the vendors to gather information related to price, platform, and additional features of the tools. I attended the conference on July 30, 1998 and interviewed 8 software vendors. The interviews consisted of questions of the vendors and demonstrations of the software.

The matrices were developed with the information gathered during the literature review, the product review and the interviews.