

First, you clean the data extracted from each source. Semantic standardization is another major task. Cleaning may just be correction of misspelling or may include resolution of conflicts between state codes and zip codes in the source data, or may deal with providing default values for missing data elements, or elimination of duplicates when you bring in the same data from multiple source systems. You standardize the data types and field lengths for same data elements retrieved from the various sources. Data transformation involves many forms of combining pieces of data from the different-sources. You resolve synonyms and homonyms