to improve the efficiency of detection. Many current commercial tools support the extraction, transformation, and loading (ETL) [2, 3] of (possibly unclean) data into a trustworthy (cleansed) database. ProbClean [2] is another solution to efficiently support relational queries. It treats duplicate detection procedures as data processing tasks with uncertain outcomes. There are many defined techniques for data cleaning, some are as follows.

2.1 Parsing

To detect the syntax errors, parsing method is used in data cleansing. By parsing, lexical errors and domain errors can be rectified as it firstly takes sample set of values to deduce the format of the domain. Besides that, for anomaly detection, it generates discrepancy detector.

2.2 Data Transformation

Data transformation is another process of data cleansing in which first of all mapping of data is done from some given format, into a common scheme, which fit it according to the needs, and then it is transformed into the format expected. Standardization and normalization are the part of transformations before the mapping to remove irregularities in data.

2.3 Integrity Constraint Enforcement

Integrity is the major concern when data is modified by inserting, deleting, or updating something. If some integrity constraints are violated, then it is rejected during integrity constraint checking. Additional identified updates are to be added only to the original data if there is no violation in integrity constraint.

2.4 Duplicate Elimination

Duplicate elimination is an essential part of data cleansing. There are a number of methods in duplicate elimination, in every method. In each duplicate detection method, there must be an algorithm which detects the duplication in each entry [4–7].