**REJECT STRATEGIES IMPLEMENTED**

**In Staging:**

1. Stg\_box\_office\_worldwide

In this table, we are considering the first level of accepted top 1000 movies records to be the ones with the matched column values of primaryTitle and startYear from stg\_imdb\_title\_basics to Title and Year column values from the top 1000 movies tsv file respectively; unmatched ones are rejected in the first level. Then, the second level of accepted top 1000 movies records is obtained by mapping the originalTitle and startYear from stg\_imdb\_title\_basics to Title and Year values from the top 1000 movies tsv file respectively; unmatched ones are rejected in the second level. Then, the third level of accepted top 1000 movies records is obtained by mapping the startYear from stg\_imdb\_title\_basics to the column Year incremented by 1 from the top 1000 movies tsv file to correct the data in Year and startYear columns in both sources; unmatched ones are rejected in the third level. Then, the fourth level of accepted top 1000 movies records is obtained by mapping the startYear from stg\_imdb\_title\_basics to the column Year decremented by 1 from the top 1000 movies tsv file to correct the data in Year and startYear columns in both sources; unmatched ones are rejected in the fourth level and this is the final level of rejects.

**In Integration:**

1. Dim\_imdb\_genres\_ml\_rejects

In this table, we are considering the rejected records to be the ones present in MovieLens dataset and not present in IMDB dataset. We are achieving this by mapping the genres column from stg\_ml\_movies to genres column in dim\_imdb\_genres and performing a join on movieId column from dim\_imdb\_title\_basics (lookup) with movieId from stg\_ml\_movies (main) and catching the rejected values by setting the “Catch lookup inner join reject” to true in the dim\_imdb\_genres\_ml\_rejects table.

1. Dim\_imdb\_name\_basics\_knownForTitles\_rejects

In this table, we are considering the rejected records to be the normalized knownForTitles column values in stg\_imdb\_name\_basics\_knownForTitles not matching with the tconst column values in dim\_imdb\_title\_basics. We are achieving this by mapping the nconst column from stg\_imdb\_name\_basics\_knownForTitles to nconst column in Dim\_imdb\_name\_basics and performing an inner join on knownForTitles column from dim\_imdb\_name\_basics\_knownForTitles with tconst from dim\_imdb\_title\_basics and catching the rejected values by setting the “Catch lookup inner join reject” to true in the m\_imdb\_name\_basics\_knownForTitles\_rejects table.

1. Dim\_imdb\_title\_akas\_rejects and Dim\_imdb\_title\_principals\_rejects

In these tables, we are obtaining the rejected records by setting the “Catch output reject” as true in the tMap of corresponding reject tables.

1. Dim\_imdb\_title\_crew\_directors\_rejects and Dim\_imdb\_title\_crew\_writers\_rejects

In these tables, we are considering the rejected records to be the ones with the unmatched tconst and nconst values from dim\_imdb\_title\_basics and dim\_imdb\_name\_basics respectively with the tconst and nconst values from stg\_imdb\_ title\_crew\_directors/ stg\_imdb\_ title\_crew\_writers tables.

1. All fct\_ml\_rejects tables

In these tables, we are considering the rejected records to be the ones with the unmatched movieId column values from the corresponding stg\_ml tables with the movieId column values from dim\_imdb\_title\_basics; an inner join is performed and the “Catch lookup inner join reject” value is set to true in the all the fct\_ml\_rejects tables.