Step 1: Leveling the keywords.csv file

Initially, the file was read and converted to pandas Dataframe for easier management. With a for loop, we ran the 'keywords' column, using ast.literal_eval() to convert the Json representation to Dicts. Then two lists, movie_keyword_pairs and id_name, were initialized to store the data. For each line, 'movie_id' is extracted and with an emulated for running through the 'keywords', 'id' and 'name' are extracted. Finally, by filling the new columns with the corresponding data, and converting them to DataFrames, , movie_keyword_df and id_name_df are created, In the second Daraframe, duplicate id values are removed using the drop_duplicates() method. Finally, the processed data is saved in separate CSV files, 'hasKeyword.csv' and 'Keyword.csv', using to_csv().

Step 2: Import Data into my database

Using SQL Server Import, the tables of the Movielens database files were imported one by one. Each file was also a table, with the exception of the keywords.csv file, which due to flattening became two files. In each table import the data type was adjusted appropriately.

Step 3: Create primitives and foreign keys

For the tables: movie, genre, productioncompany, collection, movie_cast, movie_crew and keyword, after studying their data through The Movies Dataset (kaggle.com), the columns that have unique characteristics were identified and constitute the primary keys. Then with the alter table method the primary keys were defined. In a similar way the secondary keys were defined for the tables belongsTocollection, hasGenre, hasProductionCompany, Ratings, movie_cast, movie_crew, hasKeyword.

Table 1: Description of the primary keys.

Table	Primary key column	Key name
movie	id	movie_pk
genre	id	genre_pk
productioncompany	id	productioncompany_pk
collection	id	collection_pk
movie_cast	cid	moviecast_pk
movie_crew	cid	moviecrew_pk
Keywords	id	keywords_pk

Table 2: Description of foreign keys.

Table	Foreign key column	Key name	Reference Table
ratings	movie_id	FK_movieid	movie(id)
belongsTocollection	movie_id	FK_movied	movie(id)
belongsTocollection	collection_id	FK_collectioned	collection(id)
hasGenre	genre_id	FK_genreid	genre(id)
hasGenre	movie_id	FK_genremovieid	movie(id)
hasProductioncompany	pc_id	FK_pcid	productioncompany(id)
hasProductioncompany	movie_id	FK_productionmovied	movie(id)
movie_cast	movie_id	FK_castmovieid	movie(id)
movie_crew	movie_id	FK_crewmovieid	movie(id)
hasKeyword	keyword_id	FK_keywordid	Keywords(id)