Extracting The total storage size of the IMDB datasets and the Oscar Awards dataset we used amounts to around

The group chose to use all of the remaining datasets and extracted only the columns that will be relevant for the O

Transforming For the transformation process, the group took into account null and mismatched values and transfor

Multiple columns in the datasets contain string array values separated by commas that the group needed to accoun

For string arrays that do not have a fixed amount of values, such as the person keys from the crews and Oscars tab

Since the fact tables were made with faster querying in mind, then some of its values have to be queried from the o

The group added foreign key constraints to various keys to maintain data integrity, thus the group also used the *IN*

Loading Initially, the group wanted to load the datasets into the data warehouse using Python, however its operatio

Additionally, a major problem the group encountered during the loading process was accounting for null and misma

The ETL process takes a long time to execute due to the sheer amount of data across all datasets that the script ne