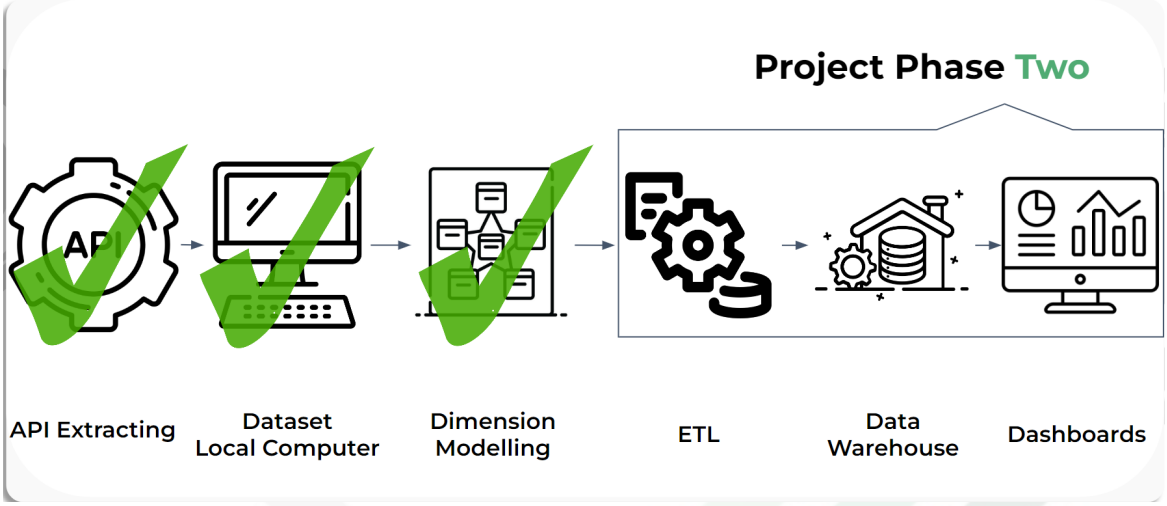
POTENTIA's summer training Project
Phase two



3. Data transformation and loading

As an expert in using python for data transformation and validation, we need your help to use your knowledge to transform the movies dataset to the appropriate shape referring to the schema you made earlier, that is, we need first to make sure that we apply first the **basic transformation** to make sure our data is clean and in the correct format, then, we apply the **advanced transformation** to enhance the quality of our data, **the more** insights you gain about the data **the more** it will be in its most clean and organized shape.

Basic transformation

1. Dealing with nulls (dropping it).
2. Format revision (data types, charset, date and time format).
3. Relationships restructuring by using primary and foreign keys.

Advanced transformation

1. Derivation (deriving new attributes)
2. Dropping the unwanted columns for analysis (as the business owner specified before)
3. Splitting (this is a useful step for dealing with multivalued attributes, splitting them into multiple single valued rows)
4. Summarization (this step is for you, It will help you gain more insights about the data so make a good usage of it)

Important notes:

Make sure to generate IDs for tables that don't have IDs (you can use NumPy arrays for this), also make sure to apply transformation on multiple tables simultaneously, each in its own .csv file.

While creating the fact table, foreign keys must be same size as its primary key (for example: 50 different primary key values => 50 different foreign keys values)

Now, our expert is ready to load the data, you are asked to load the data on **SQLServer** and construct the needed relationships.

4. Reporting using Power BI

Congratulations! You have proven yourself as a data engineer expert.
The manager's aim right now after gathering the data in the DWH is to make our movies receive the "Must-See" award , and to receive it the movie's Metascore must be 81 or higher . She asked you to tell her the average metascore for each genre and if the release time affects the metascore. Moreover she needs deep analysis to be done on the movies with a metascore higher than or equal to 81 and she wants to know your interpretation for this analysis. She needs this report to be ready as soon as possible.

Take into consideration :

- Your dashboard must be given a meaningful title.

- Your dashboard should contain:
  - A chart showing the average metascore of each genre.
  - A graph showing the effect of release time on the metascore of the movie's.
  - Deep analysis on the movies with metascore higher than or equal to 81 (e.g. find the factors affecting the metascore)
  - Don't forget the call to action.

- You must use at least 3 different types of visualizations in your dashboard with 2 different slicers.
- Feel free to get creative with more additions