ETL PROJECT_API_1
Pentaho Data Integration

Pentaho Transformations

Pentaho Data Integration(PDI) provides the Extract, Transform, and Load (ETL) capabilities that facilitate the process of capturing, cleansing, and storing data using a uniform and consistent format that is accessible and relevant to end users and IoT technologies.

This is a 5 steps ETL process, this transformation extract data from a site in a JSON format, clean, correct, add or remove…etc then loaded to sql Database.

1. Extraction. Generate rows will help to get and read the site url and fill the extracting data type.
2. Rest Client. The Rest Client will read the url, choose the file format then name it.
3. Json Input. This is will finally extract from the root path, define its data type.
4. Select Value. This is the transformation and last data preparation before the Load happen.
5. Table Output. Finally, it's the creation of the landing table in a sql database for further analysis.

By running the transformation, a green tick will appear if everything is correct as pictured above.