The system consists of 19 tasks

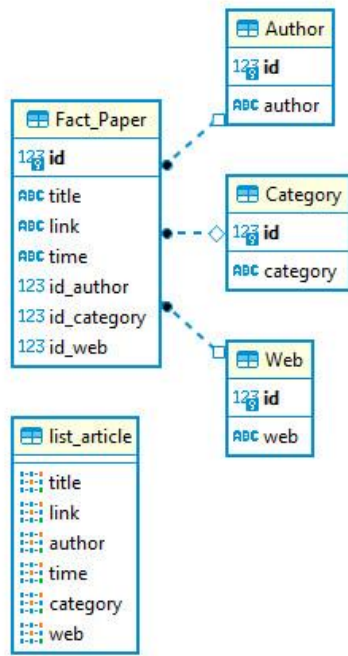The first task is to create a temporary table to store the latest articles

The next 4 tasks are to see which newspapers are new and which are their

own

Next 4 tasks After receiving the new notification, we start downloading data

Next 4 tasks Then upload data to temporary storage data lake

Then proceed to check whether there is a new newspaper or not, if there is a new newspaper, return the create_data_main task, otherwise return the no_new_newspaper task, send the report and finish.

Main storage table

After creating the next table will ETL data to the database for storage, the database will store all downloaded articles.

Then use regex to analyze and email the report

To launch the project, visit the link: http://localhost:8080/