**Setup/Installation guide:**

**Database setup:**

As, I have used remote database of Oracle 18C. To make connection and use this database, we need to follow below steps:

* pip install cx\_Oracle

Install Cx\_Oracle client on windows. For python 3.5 and above:

* python -m pip install cx\_Oracle –upgrade
* Download client with the link: <https://www.oracle.com/database/technologies/instant-client.html>
* Extract client downloaded file
* Copy all .dll files to the root python folder

Great! Now you will be able to get connected with the oracle database for use.

As we have worked on files system, we have not used much used the external APIs. We need to install few. You can also install them by following below commands:

* py -m pip install pandas
* py -m pip install os
* py -m pip install dask.dataframe
* py -m pip install json

Create a folder inside the project folder named: “Dataset” and subfolders named: “Evidence”, “Target”, “Disease” and download all the respective file given in the repository with the link below:

For Evidence: <http://ftp.ebi.ac.uk/pub/databases/opentargets/platform/21.11/output/etl/json/evidence/sourceId%3Deva/>

For Target: <http://ftp.ebi.ac.uk/pub/databases/opentargets/platform/21.11/output/etl/json/targets/>

For Disease: <http://ftp.ebi.ac.uk/pub/databases/opentargets/platform/21.11/output/etl/json/diseases/>

Now you will be able to test the given code.