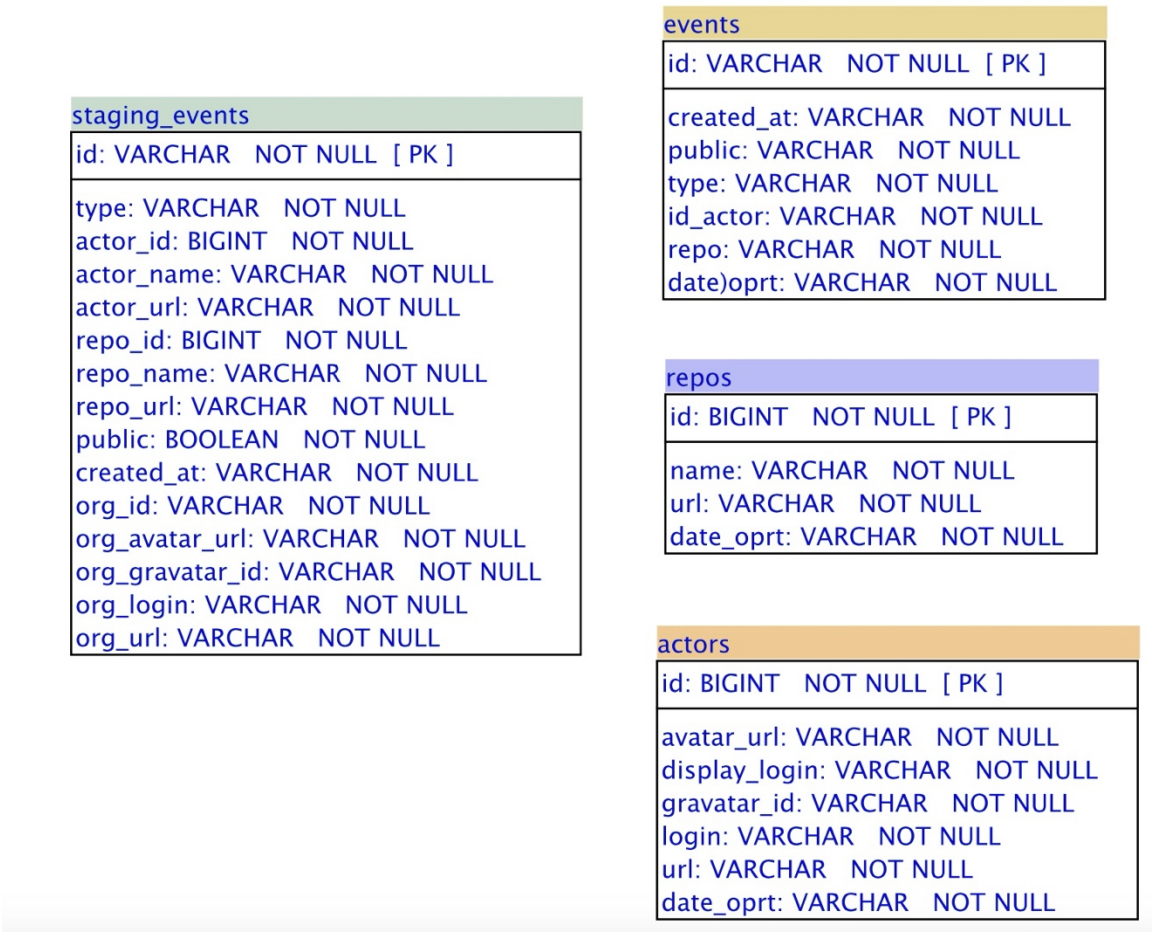
Week 4 : 04-building-a-data-lake

Data model



Project Processing

1. change directory to project 04-building-a-data-lake:

   $ cd 04-building-a-data-lake

2. Applying code for saving jupyter lab (Any update on coding)

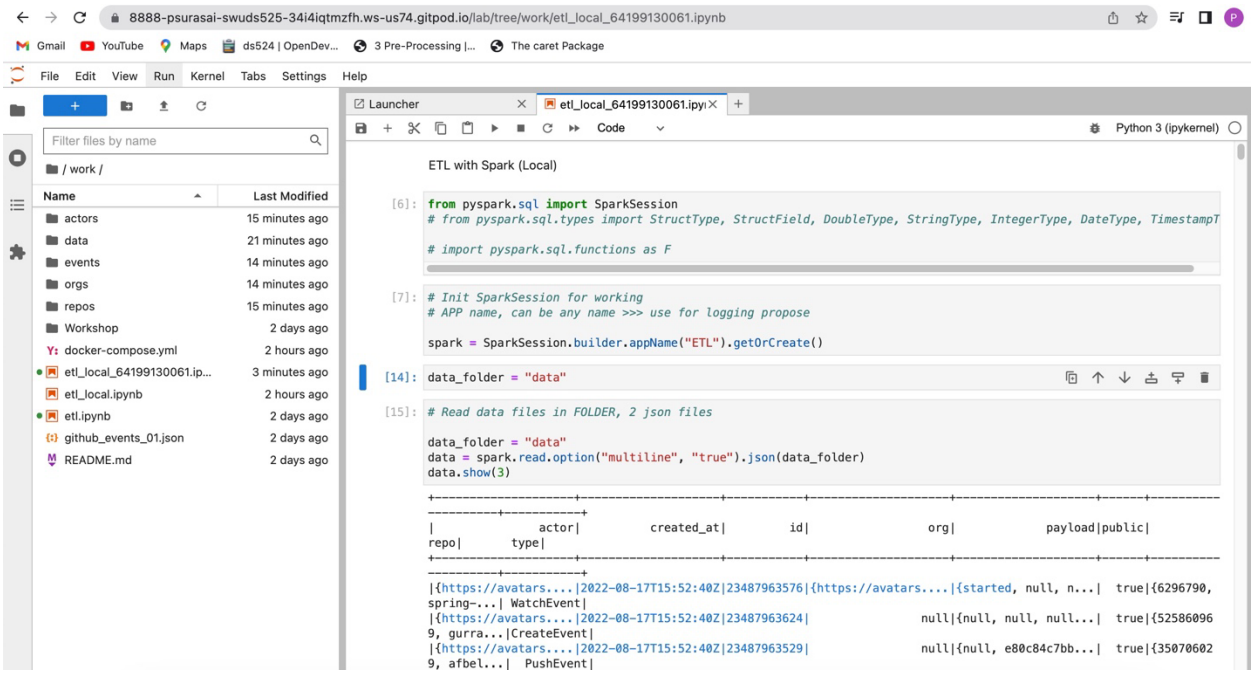
   sudo chmod 777 .

3. prepare environment workspace by Docker:

   $ docker-compose up

4. Open JupyterLab URL:

5. Execute the Notebook ‘etl_local_64199130061.ipynb' step by step:



6. Chcek the cleaned output data in folders, partition by 'date_oprt':

actors : [actors](https://github.com/psurasai/SWU-DS525/tree/main/04-building-a-data-lake/actors)

repos : [repos](https://github.com/psurasai/SWU-DS525/tree/main/04-building-a-data-lake/repos)

orgs : [orgs](https://github.com/psurasai/SWU-DS525/tree/main/04-building-a-data-lake/orgs)

events : [events](https://github.com/psurasai/SWU-DS525/tree/main/04-building-a-data-lake/events)

7. Shutdown environment workspace:

   $ docker-compose down