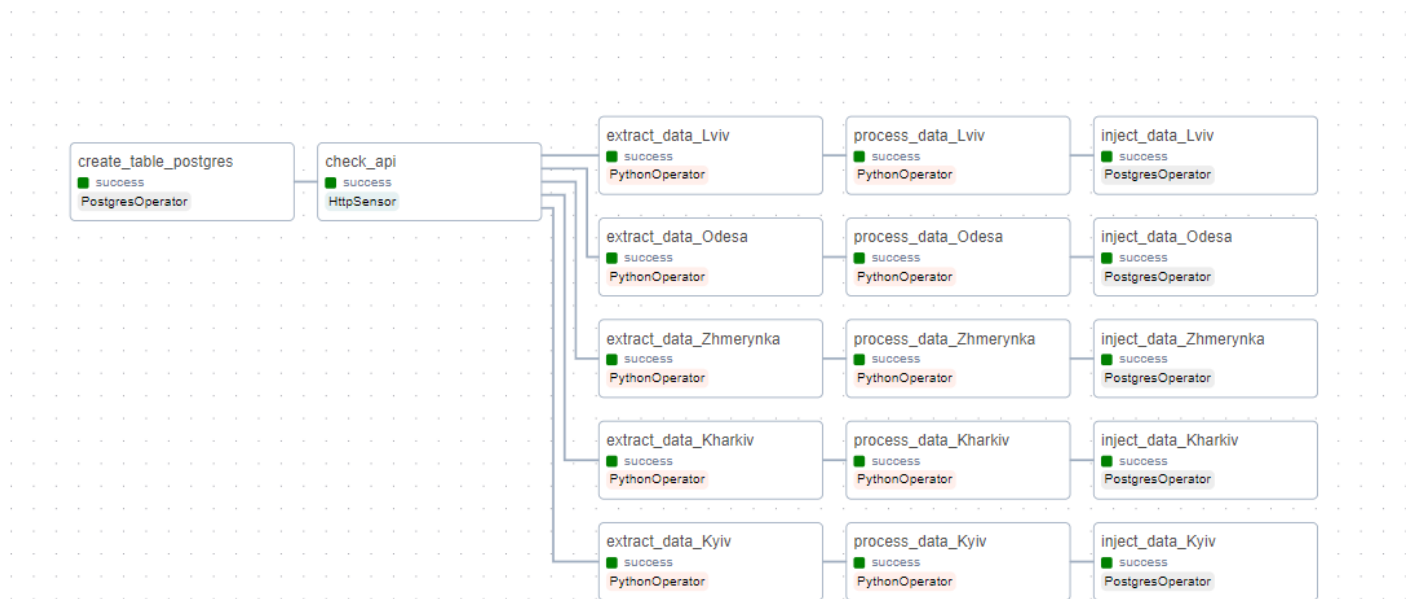


I use PostgreSQL for airflow db and also use PostgreSQL db for saving weather data.  
Here is how DAG looks:



Jobs:

### **create\_table\_postgres**

This job creates table in postgres with columns city, timestamp, date, temp, humidity, clouds, wind\_speed

### **check\_api**

This job checks whether OpenWeather API is working

### **extract\_data\_{city}**

This job extracts data from OpenWeather for job DAG start date and for given city

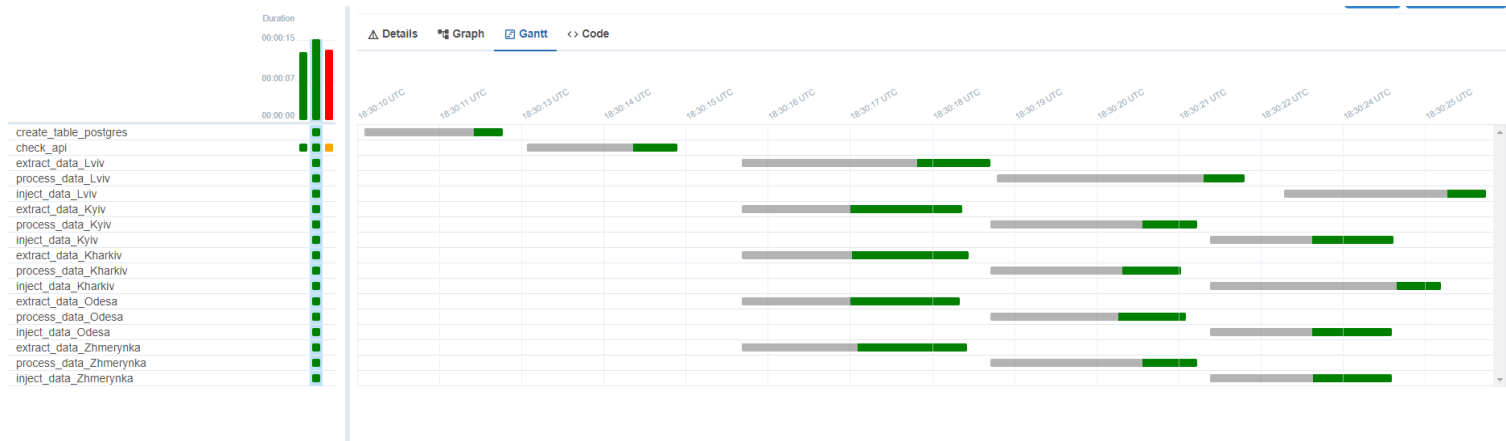
### **process\_data\_{city}**

This jobs processes data for given city and returns only values that need to be injected in table

### **inject\_data\_{city}**

This job injects data into previously created PostgreSQL table

Jobs **extract\_data\_{city}**, **process\_data\_{city}** and **inject\_data\_{city}** are running in parallel  
It can be seen on Gantt chart:



Here is the results of running pipeline for two different dates:

weather_db=#	select * from measures;						
city	timestamp	date	temp	humidity	clouds	wind_speed	
Kharkiv	2023-12-01 20:29:09	2023-12-01 18:29:09	1.85	98	100	3.4	
Zhmerynka	2023-12-01 20:29:09	2023-12-01 18:29:09	-0.33	98	100	2.95	
Odesa	2023-12-01 20:29:09	2023-12-01 18:29:09	9.96	77	96	3.21	
Lviv	2023-12-01 20:29:09	2023-12-01 18:29:09	-0.64	99	100	1.91	
Kyiv	2023-12-01 20:29:09	2023-12-01 18:29:09	0.19	92	100	1.86	
Odesa	2023-11-11 02:00:00	2023-11-11 00:00:00	14.55	81	100	14.23	
Zhmerynka	2023-11-11 02:00:00	2023-11-11 00:00:00	9.31	71	100	8.46	
Kyiv	2023-11-11 02:00:00	2023-11-11 00:00:00	7.53	81	100	0.89	
Kharkiv	2023-11-11 02:00:00	2023-11-11 00:00:00	4.55	85	100	3.55	
Lviv	2023-11-11 02:00:00	2023-11-11 00:00:00	5.71	77	100	2.69	
(10 rows)							