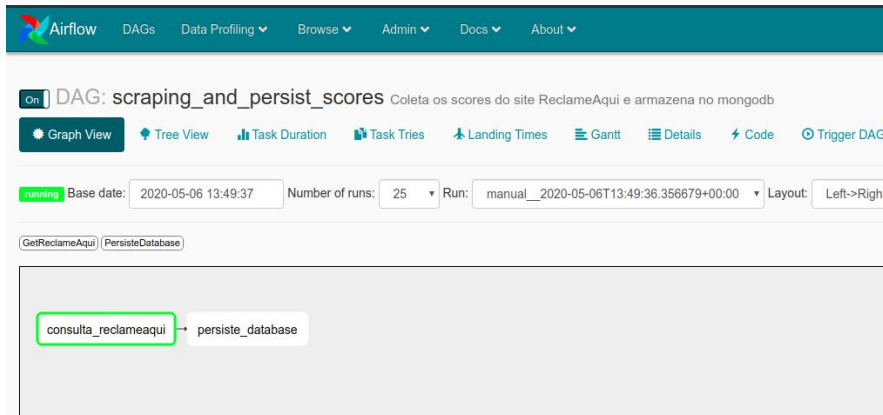


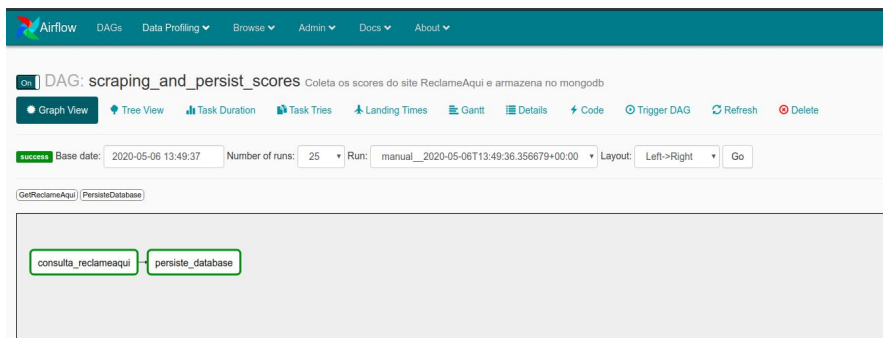
1) Na tela inicial do Airflow, clicar no botão de play do Job

	DAG	Schedule	Owner	Recent Tasks	Last Run	DAG Runs	
	On scraping_and_persist_reclamations	1 day, 6:00:00	airflow		2020-05-05 00:00		
	On scraping_and_persist_scores	1 day, 6:00:00	airflow		2020-05-06 13:45		

2) Click na data (Last Run) do Job e abra a DAG de execução



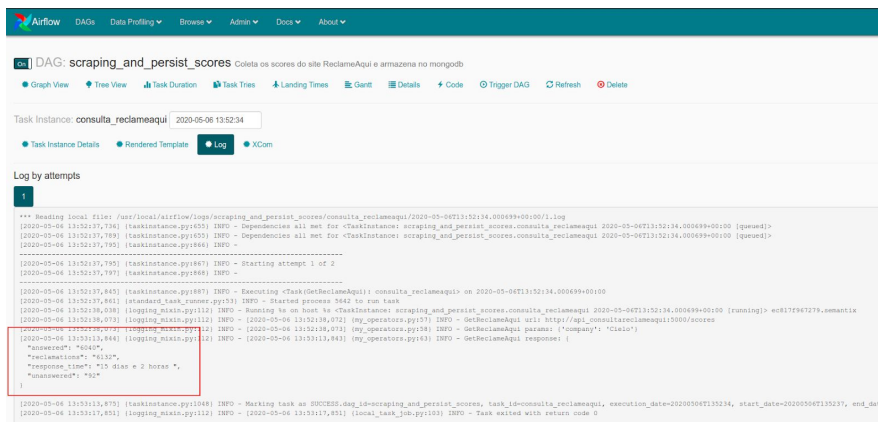
3) Aguarde a execução de todas as task. (o status “sucess” irá aparecer no final)



4) Click na primeira caixinha “consulta\_reclameaqui” e em “View log”

The screenshot shows the Airflow task instance details for 'consulta\_reclameaqui' on 2020-05-06T13:52:34.000699+00:00. The 'View Log' button is highlighted with a red circle. Below the buttons, there is a 'Download Log (by attempts):' section with a '1' attempt listed. The 'Run' button is also visible.

## 5) As informações coletadas pelo scraping deverão estar visíveis

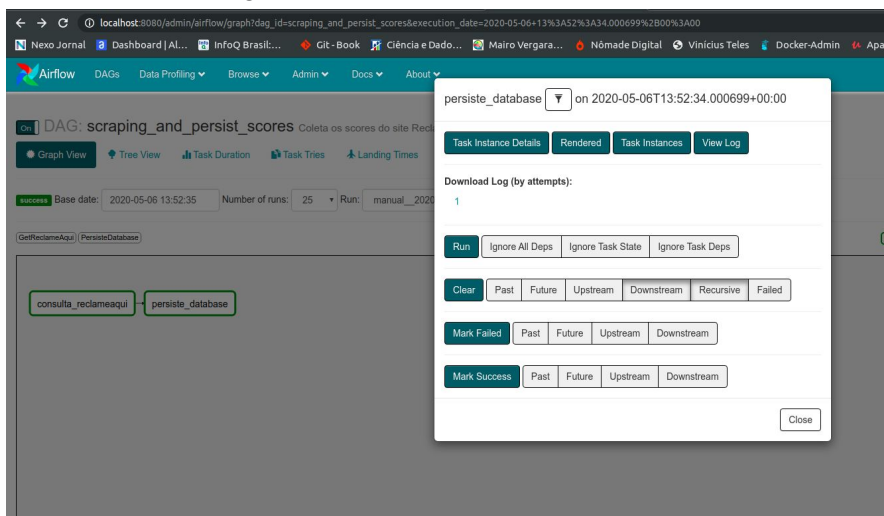


Task Instance: `consulta_reclameaqui` 2020-05-06 13:52:34

Log by attempts

```
1
*** Reading local file: /usr/local/airflow/logs/scraping_and_persist_scores/consulta_reclameaqui/2020-05-06T13:52:34.000699+00:00/1.log
[2020-05-06 13:52:37,786] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.consulta_reclameaqui 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,789] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.consulta_reclameaqui 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,795] [taskinstance.py:1666] INFO -
-----
[2020-05-06 13:52:37,795] [taskinstance.py:1667] INFO - Starting attempt 1 of 2
[2020-05-06 13:52:37,797] [taskinstance.py:1668] INFO -
-----
[2020-05-06 13:52:37,845] [taskinstance.py:1687] INFO - Executing <Task(GetReclameAqui): consulta_reclameaqui> on 2020-05-06T13:52:34.000699+00:00
[2020-05-06 13:52:37,845] [standalone_task_runner.py:33] INFO - Started process 3642 to run task
[2020-05-06 13:52:38,038] [logging_mixin.py:112] INFO - Running task on host <TaskInstance: scraping_and_persist_scores.consulta_reclameaqui 2020-05-06T13:52:34.000699+00:00 [running]> ec8176967278.semanti
[2020-05-06 13:52:38,073] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,073) [my_operators.py:157] INFO - GetReclameAqui url http://api.consultareclameapi15000/scores
[2020-05-06 13:52:38,073] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,073) [my_operators.py:161] INFO - GetReclameAqui params: {'company': 'Cielo'}
[2020-05-06 13:52:38,644] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,644) [my_operators.py:163] INFO - GetReclameAqui response: {
  "answered": "6040",
  "reclamations": "6152",
  "response_time": "15 dias e 2 horas ",
  "unanswered": "32"
}
[2020-05-06 13:52:38,675] [taskinstance.py:11048] INFO - Marking task as SUCCESS: dag_id=scraping_and_persist_scores, task_id=consulta_reclameaqui, execution_date=20200506T135234, start_date=20200506T135237, end_date=
[2020-05-06 13:52:37,851] [logging_mixin.py:112] INFO - (2020-05-06 13:52:37,851) [local_task_job.py:103] INFO - Task exited with return code 0
```

## 6) Click na segunda caixinha “persiste\_database” e em “View log”



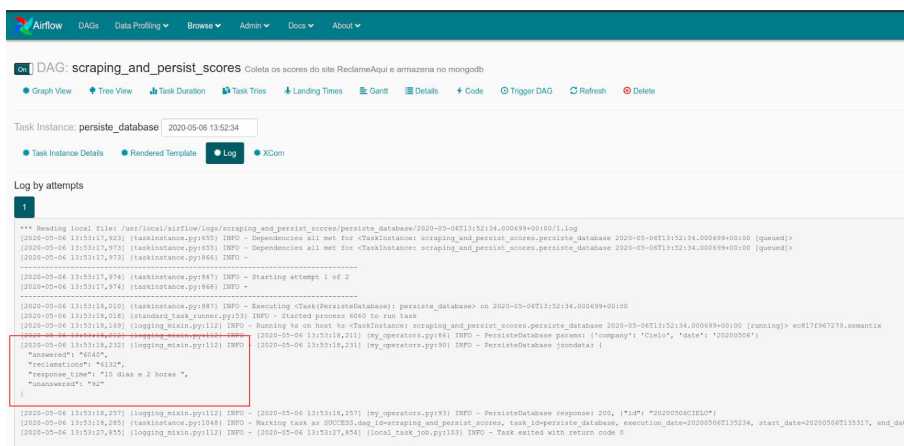
Task Instance: `persiste_database` on 2020-05-06T13:52:34.000699+00:00

Task Instance Details | Rendered | Task Instances | View Log

Download Log (by attempt):

```
1
[2020-05-06 13:52:37,845] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,845] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,851] [taskinstance.py:1666] INFO -
-----
[2020-05-06 13:52:37,851] [taskinstance.py:1667] INFO - Starting attempt 1 of 2
[2020-05-06 13:52:37,851] [taskinstance.py:1668] INFO -
-----
[2020-05-06 13:52:38,038] [taskinstance.py:1687] INFO - Executing <Task(PersistDatabase): persiste_database> on 2020-05-06T13:52:34.000699+00:00
[2020-05-06 13:52:38,038] [standalone_task_runner.py:33] INFO - Started process 3640 to run task
[2020-05-06 13:52:38,149] [logging_mixin.py:112] INFO - Running task on host <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [running]> ec8176967278.semanti
[2020-05-06 13:52:38,182] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,182) [my_operators.py:164] INFO - PersistDatabase params: {'company': 'Cielo', 'date': '20200506'}
[2020-05-06 13:52:38,211] [my_operators.py:161] INFO - PersistDatabase params: {'company': 'Cielo', 'date': '20200506'}
[2020-05-06 13:52:38,211] [my_operators.py:163] INFO - PersistDatabase response: {
  "answered": "6040",
  "reclamations": "6152",
  "response_time": "15 dias e 2 horas ",
  "unanswered": "32"
}
[2020-05-06 13:52:38,257] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,257) [my_operators.py:161] INFO - PersistDatabase response: 200, {'data': '20200506T135237'}
[2020-05-06 13:52:38,285] [taskinstance.py:11048] INFO - Marking task as SUCCESS: dag_id=scraping_and_persist_scores, task_id=persiste_database, execution_date=20200506T135234, start_date=20200506T135317, end_date=
[2020-05-06 13:52:37,855] [logging_mixin.py:112] INFO - (2020-05-06 13:52:37,855) [local_task_job.py:103] INFO - Task exited with return code 0
```

## 7) As informações armazenadas no mongoDB deverão estar visíveis e serem as mesmas coletadas pela task anterior



Task Instance: `persiste_database` 2020-05-06 13:52:34

Log by attempts

```
1
*** Reading local file: /usr/local/airflow/logs/scraping_and_persist_scores/persiste_database/2020-05-06T13:52:34.000699+00:00/1.log
[2020-05-06 13:52:37,823] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,873] [taskinstance.py:1655] INFO - Dependencies all met for <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [queued]>
[2020-05-06 13:52:37,873] [taskinstance.py:1666] INFO -
-----
[2020-05-06 13:52:37,874] [taskinstance.py:1667] INFO - Starting attempt 1 of 2
[2020-05-06 13:52:37,874] [taskinstance.py:1668] INFO -
-----
[2020-05-06 13:52:38,038] [taskinstance.py:1687] INFO - Executing <Task(PersistDatabase): persiste_database> on 2020-05-06T13:52:34.000699+00:00
[2020-05-06 13:52:38,038] [standalone_task_runner.py:33] INFO - Started process 3640 to run task
[2020-05-06 13:52:38,149] [logging_mixin.py:112] INFO - Running task on host <TaskInstance: scraping_and_persist_scores.persiste_database 2020-05-06T13:52:34.000699+00:00 [running]> ec8176967278.semanti
[2020-05-06 13:52:38,182] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,182) [my_operators.py:164] INFO - PersistDatabase params: {'company': 'Cielo', 'date': '20200506'}
[2020-05-06 13:52:38,211] [my_operators.py:161] INFO - PersistDatabase params: {'company': 'Cielo', 'date': '20200506'}
[2020-05-06 13:52:38,211] [my_operators.py:163] INFO - PersistDatabase response: {
  "answered": "6040",
  "reclamations": "6152",
  "response_time": "15 dias e 2 horas ",
  "unanswered": "32"
}
[2020-05-06 13:52:38,257] [logging_mixin.py:112] INFO - (2020-05-06 13:52:38,257) [my_operators.py:161] INFO - PersistDatabase response: 200, {'data': '20200506T135237'}
[2020-05-06 13:52:38,285] [taskinstance.py:11048] INFO - Marking task as SUCCESS: dag_id=scraping_and_persist_scores, task_id=persiste_database, execution_date=20200506T135234, start_date=20200506T135317, end_date=
[2020-05-06 13:52:37,855] [logging_mixin.py:112] INFO - (2020-05-06 13:52:37,855) [local_task_job.py:103] INFO - Task exited with return code 0
```