

Введение Airflow

Дмитрий Браженко

Зачем?

1. Обработка сложных workflow
2. Отслеживание регулярных запусков
3. Отслеживание зависимостей в workflow

Зачем нужен Airflow если есть CRON

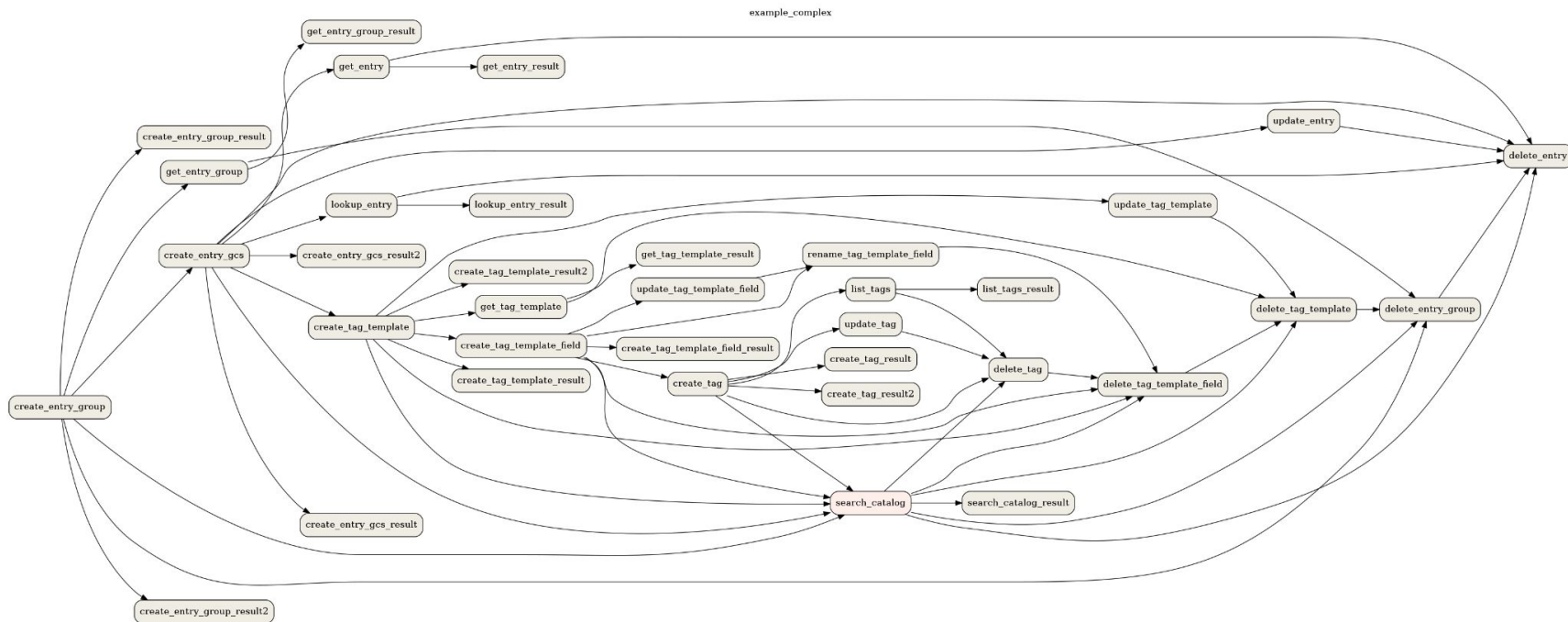
```
root@utls-newsletter-php7:~ # crontab -l
# do daily/weekly/monthly maintenance
# min    hour    day    month    weekday  command
*/15     *      *      *      *        run-parts /etc/periodic/15min
0        *      *      *      *        run-parts /etc/periodic/hourly
0        2      *      *      *        run-parts /etc/periodic/daily
0        3      *      *      6        run-parts /etc/periodic/weekly
0        5      1      *      *        run-parts /etc/periodic/monthly

22 0 * * * "/root/.acme.sh"/acme.sh --cron --home "/root/.acme.sh" > /dev/null
*/5 * * * * curl -L -s https://newsletter.cyberciti.biz/scheduled.php > /dev/null 2>&1
root@utls-newsletter-php7:~ #
```

Настройка Airflow

- Продакшен настройка – тема для отдельной лекции :)
- В моем примере – запускается докер-образ, с которым мы взаимодействуем. После окончания вебинара будет выложен на Github

Сложные направленные графы (DAG)




Альтернативы Airflow

- Luigi
- [Яндекс.Нирвана](#) (только в Яндексе)
- Другие

Преимущества Airflow

- Удобный Web UI
- Масштабируется
- Поддерживается сообществом

И так, из чего состоит AirFlow?

 Airflow

DAGs

Security

Browse

Admin

Docs

21:11 UTC



















































































































































































































RH

DAGs

All 26Active 10Paused 16

Filter DAGs by tag

Search DAGs

DAG	Owner	Runs	Schedule	Last Run	Recent Tasks	Actions	Links
 example_bash_operator example example2	airflow	  	0 0 ***	2020-10-26, 21:08:11	             	   ...	
 example_branch_dop_operator_v3 example	airflow	  	* / 1 ***		             	   ...	
 example_branch_operator example example2	airflow	  	@daily	2020-10-23, 14:09:17	             	   ...	
 example_complex example example2 example3	airflow	  	None	2020-10-26, 21:08:04	             	   ...	
 example_external_task_marker_child	airflow	  	None	2020-10-26, 21:07:33	             	   ...	
 example_external_task_marker_parent	airflow	  	None	2020-10-26, 21:08:34	             	   ...	
 example_kubernetes_executor example example2	airflow	  	None		             	   ...	
 example_kubernetes_executor_config example3	airflow	  	None	2020-10-26, 21:07:40	             	   ...	
 example_nested_branch_dag example	airflow	  	@daily	2020-10-26, 21:07:37	             	   ...	
 example_passing_params_via_test_command example	airflow	  	* / 1 ***		             	   ...	

Недостатки Airflow

- Сложный и негибкий инструмент

Подробнее [тут](#)

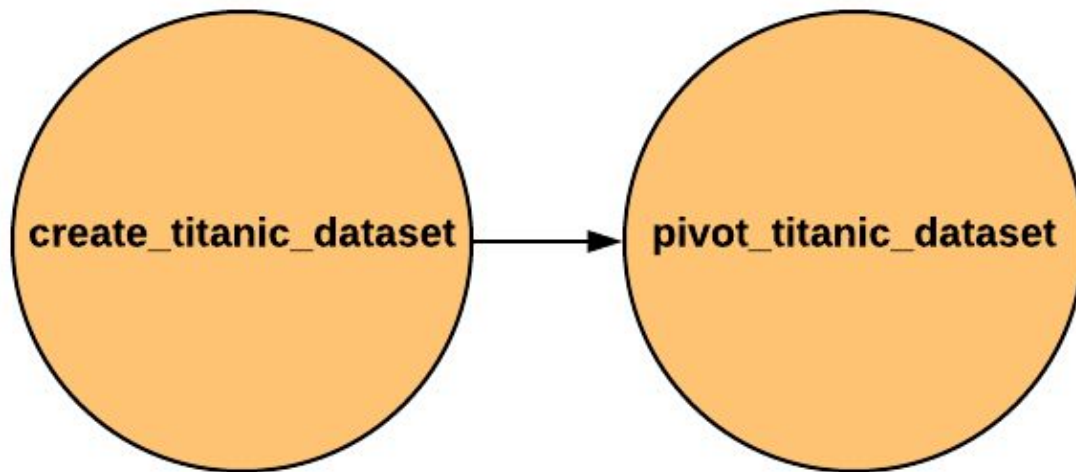
И так, из чего состоит AirFlow?

- Веб-приложение (на Flask)
- Планировщик
- Воркер, выполняющий работу
- База данных

Подробнее можно прочитать [тут](#)

Из чего состоит граф (DAG)

- Характеристик всего графа
- "Операторов" – исполняемых элементов



Построим первый DAG

1. Распечатаем текущую директорию с помощью bash команды `pwd`
2. Добавим еще один блок (`sleep`)

Построим второй DAG

Будем использовать Python операторы

Скачаем и обработаем [датафрейм](#)

Поработаем с параметрами DAG

Параметры подробно описаны в [документации](#)

Попробуем использовать их:

- Изменим расписание
- Добавим документацию к DAG

Переменные Airflow

Иногда требуется добавить переменные, которые являются общими для всех DAG'ов

[Документация](#) про хранение переменных

DockerOperator

1. Про запуск "докера в докере": [тут](#)
2. Пример с [докер операцией](#)
3. Формальная [документация](#)