



Apache
Airflow

AirFlow = workflow
management system

History

Проект начался в 2014 в Airbnb

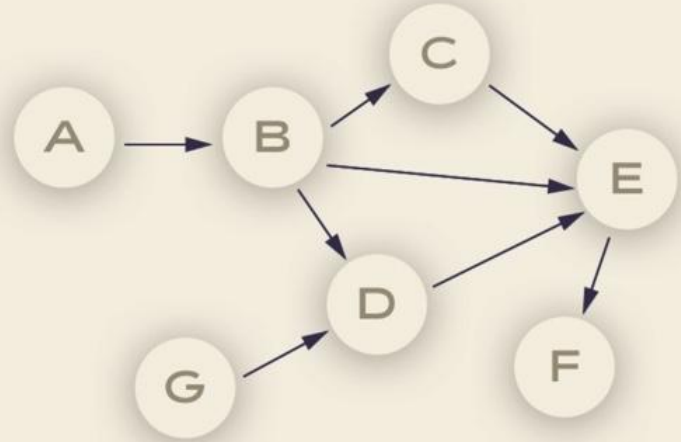
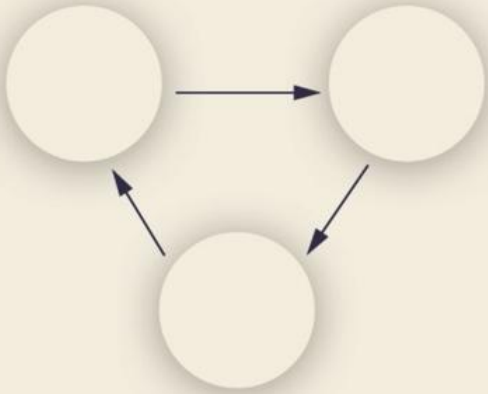
Тот же парень, что сделал superset (Maxime Beauchemin)

Open-source since 2015

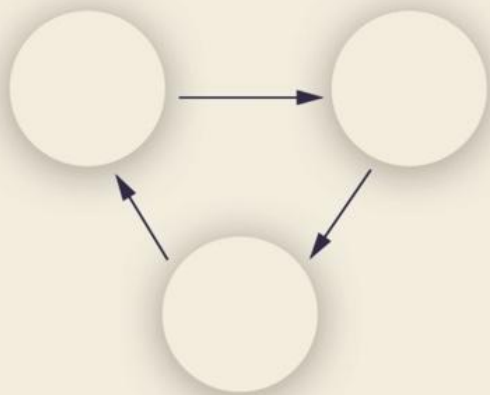
Apache incubation program March 2016

Top-Level project since Jan 2019

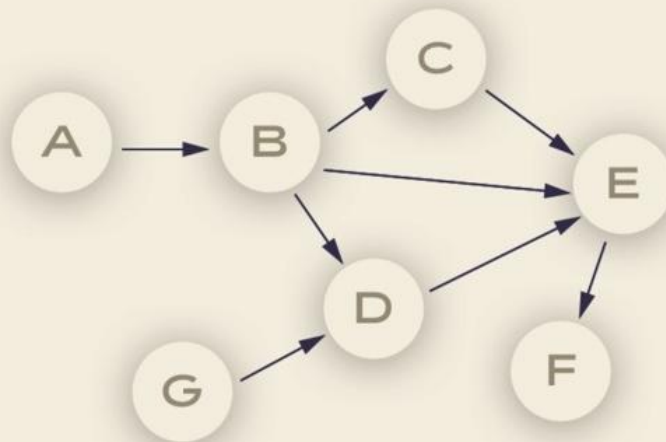
DAG (Directed Acyclic Graph)



DAG (Directed Acyclic Graph)

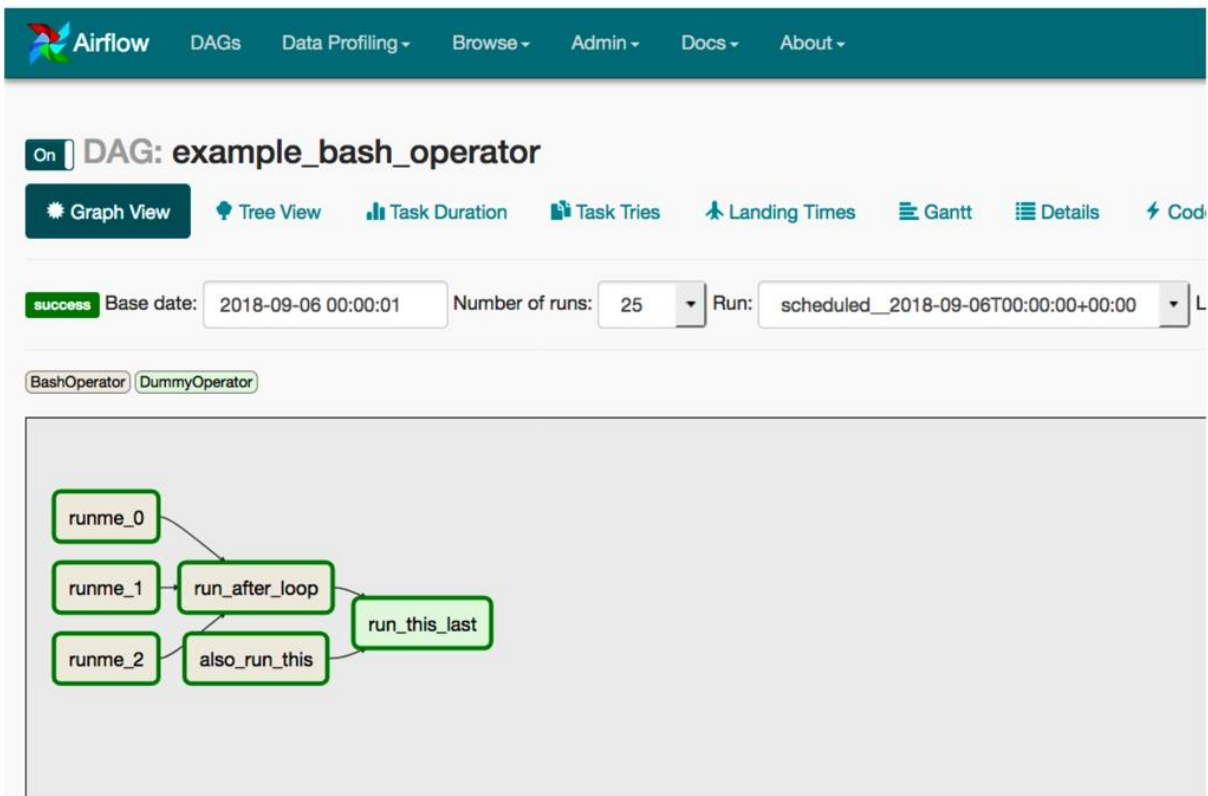


(Это не dag)



А это – dag

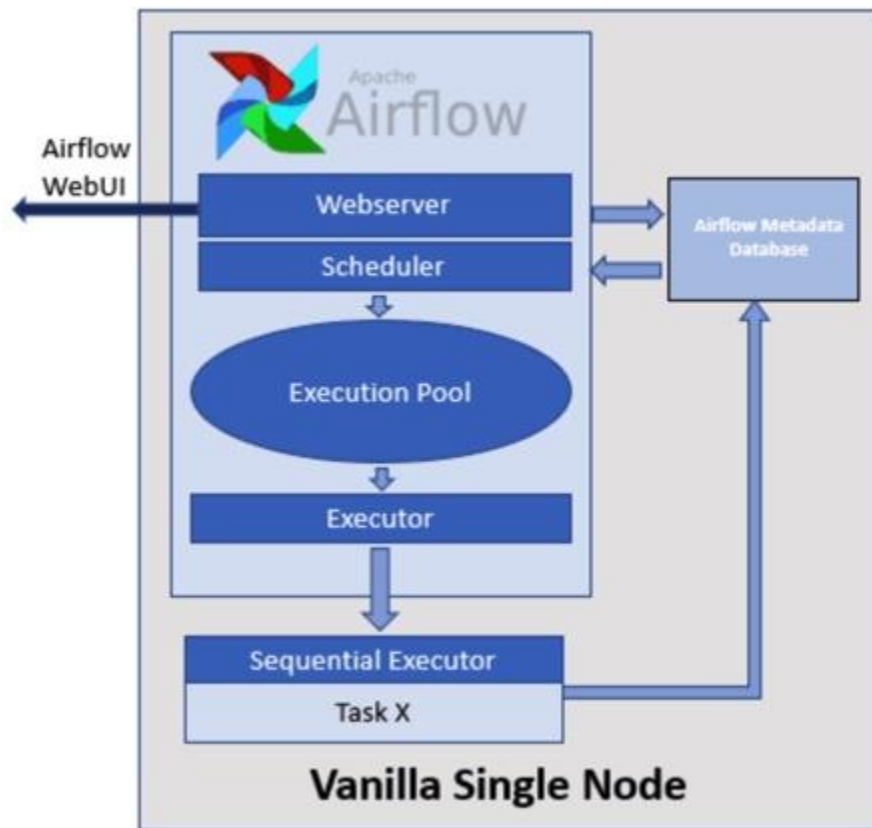
DAG (Directed Acyclic Graph)



Operator (Звено в цепочке задач)

- PythonOperator — оператор для исполнения python кода
- BashOperator — оператор для запуска bash скриптов/команд
- PostgresOperator — оператор для вызова SQL запросов в PostgreSQL БД
- RedshiftToS3Transfer — оператор для запуска UNLOAD команды из Redshift в S3
- EmailOperator — оператор для отправки электронных писем

[Полный список](#) операторов можно найти в официальной документации.



Executor

- SequentialExecutor (по умолчанию, не умеет запускать параллельные задачи)
- LocalExecutor (умеет запускать задачи параллельно, нет отказоустойчивости)
- CeleryExecutor (использует task-manager celery + брокер сообщений Redis / RabbitMQ)
- KubernetesExecutor (задачи исполняются как новые pod инстансы)