

Airflow

Применение Airflow

Skillbox

образовательная платформа

Цель видео

Рассмотреть применение Airflow
в промышленных задачах.

ETL

```
connection_list = lv.get('connection_list')
export_profiles_sql = '''
SELECT
    id,
    user_id,
    nickname,
    gender,
    {{params.shard_id}} as shard_id
FROM profiles
'''
```

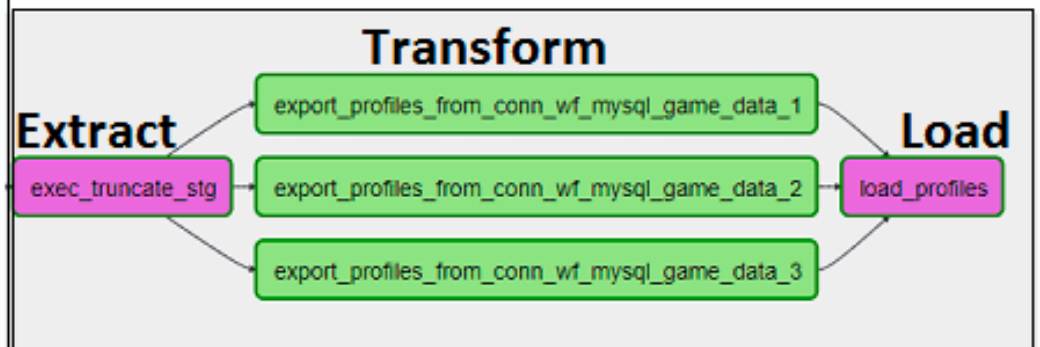
Extract

```
for conn_id in connection_list:
    export_profiles = SqlToHiveViaHdfsTransfer(
        task_id='export_profiles_from_' + conn_id,
        sql=export_profiles_sql,
        hive_table='stg.profiles',
        overwrite=False,
        tmpdir='/data/tmp',
        conn_id=conn_id,
        params={'shard_id': conn_id[-1:], },
        compress=None,
        dag=dag
    )
```

Transform

```
export_profiles.set_upstream(exec_truncate_stg)
export_profiles.set_downstream(load_profiles)
```

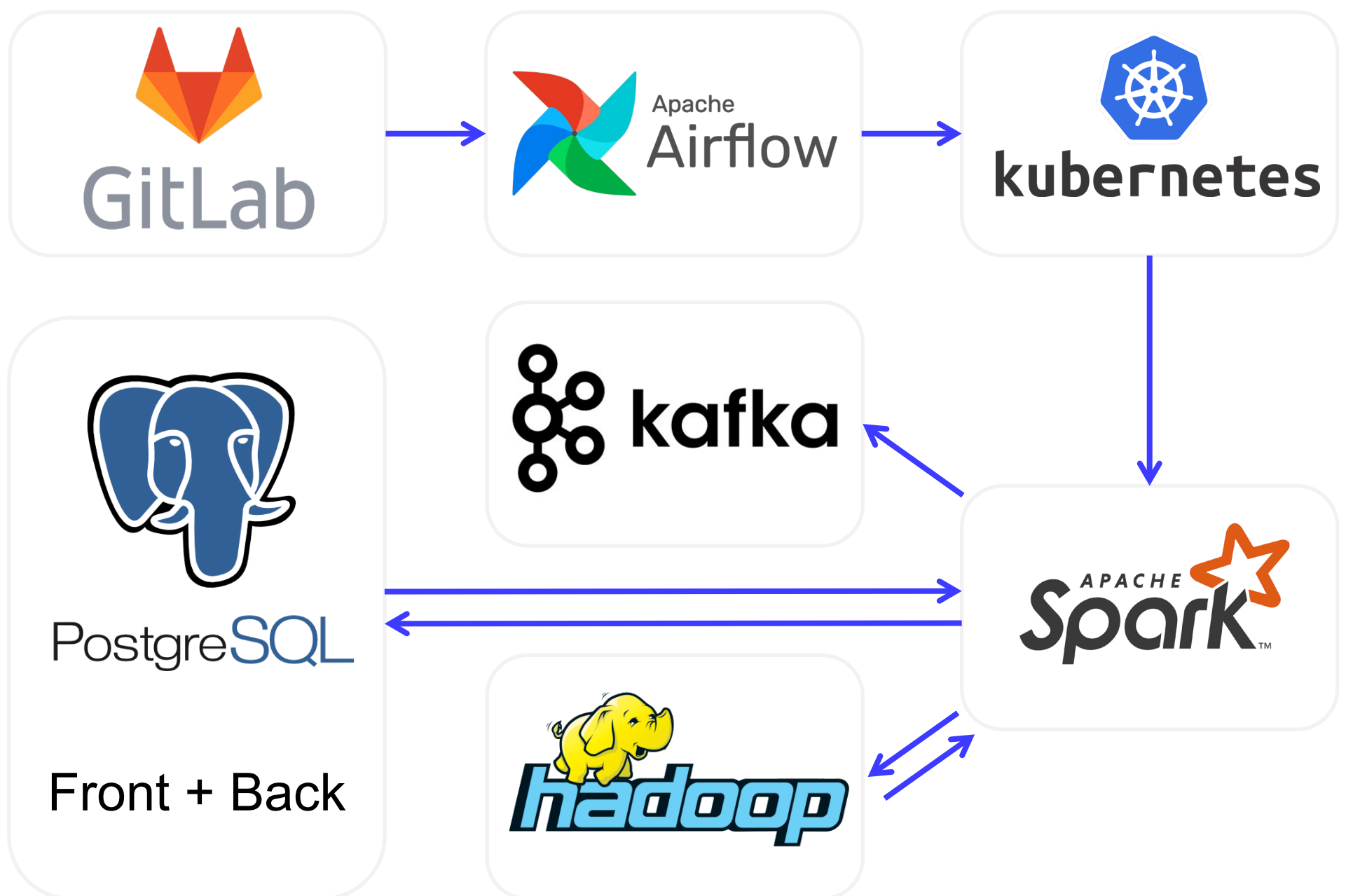
DAG



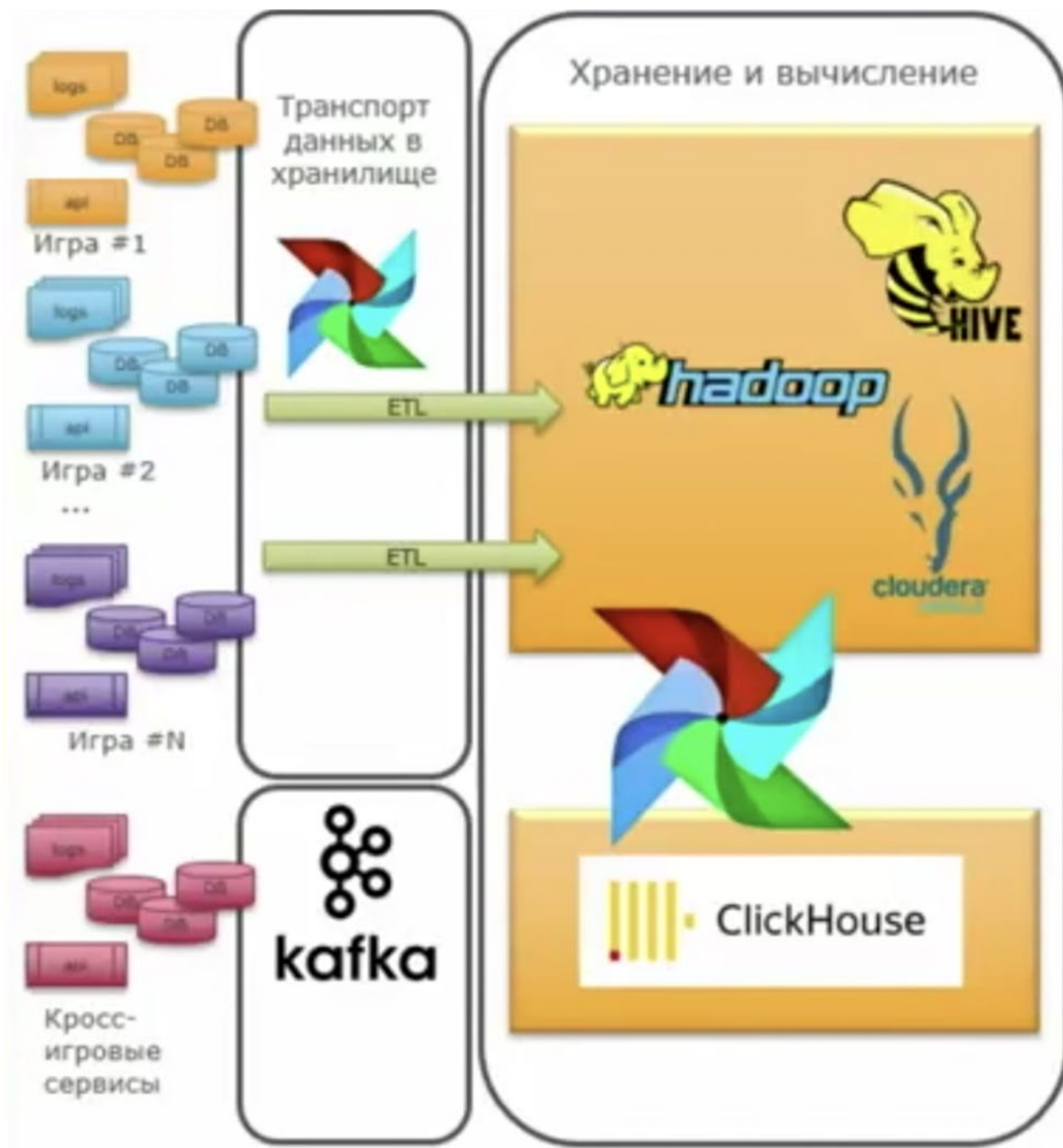
Load



Production (1)



Production (1)



Выводы

- 1 Обсудили применения Airflow в задачах дата-инженера
- 2 Разобрали две архитектуры с применением Airflow в промышленных задачах

Вывод модуля

- 1 Разобрали цели и задачи Airflow, причины его использования
- 2 Изучили основные понятия Airflow — DAG, scheduler, operator, task
- 3 Рассмотрели архитектуру Airflow
- 4 Запустили Airflow, освоили интерфейс
- 5 Настроили и запустили DAG
- 6 Разобрали настройку расписания и порядка выполнения тасок
- 7 Рассмотрели применение Airflow в промышленных масштабах