



# Продуктивизация ML- пайплайнов

вопросы и ответы на кейсах М.Видео-Эльдорадо

Карев Виталий

Разработчик Центра компетенций по  
техническим решениям для  
машинного обучения, Группа  
М.Видео-Эльдорадо

# ГРУППА М.ВИДЕО-ЭЛЬДОРАДО ЭТО...

М.ВидеоЭльдорадо

**Топ-10** 

**Крупнейшие ритейлеры  
электроники в мире**

~100m ежемесячный трафик<sup>2</sup>

**#1** 

**Онлайн-ритейлер  
электроники в России<sup>3</sup>**

33% доля рынка онлайн<sup>3</sup>

**Ведущий** 

**Ритейлер электроники в  
России<sup>3</sup>**

27% доля рынка в 2020<sup>3</sup>

**19m** 

**Активных клиентов<sup>4</sup>**

**72m база лояльных  
клиентов<sup>5</sup>**

**~60%** 

**Доля онлайн-продаж в  
обороте**

**~1m m<sup>2</sup>** 

**Площадь фулфилмент-  
центров<sup>6</sup>**

**100%** 

**Узнавание брендов<sup>7</sup>**

**RUB 0.5trln** 

**GMV<sup>9</sup>**

+15% год-к-году в 2020

**~5-7%** 

**Показатель  
маржинальности по EBITDA**

(1) Across global CE retailer specialists by revenue in 2019, data from Thomson Reuters; (2) Average monthly traffic for website, app and stores in FY 2020; (3) Data for FY 2020 market share, GfK; (4) Identified active customer denotes identified customers with at least one purchase for the last 12 months, data as of 31 Dec 2020; (5) Loyal customers represent a group of authorized customers with active loyalty cards; (6) Warehouse area in stores of ~540k m<sup>2</sup>, distribution centers of ~418k m<sup>2</sup>, and regional distribution centers of ~83k m<sup>2</sup> as of 31 Dec 2020; (7) Brand Health Tracking, 260 cities of presence M.Video and Eldorado stores (100k+ population), base 2020, customers CE during the last 6 months (9,599 respondents, Male / Female, 16-65 years), Millward Brawn A/R M/I-Marketing; (8) Q1 data for 2020, base 39/35 M.Video / Eldorado stores, survey at the exit from the store "M&P ANALYTICS LLC"; (9) Gross Merchandise Value (GMV) includes purchases in retail stores (including pick up orders), paid and delivered online orders, paid shipments to legal entities from warehouses. Purchases made in stores and online orders may be made by individuals and legal entities. GMV includes sales of goods and services, which may be owned by the Company or agents. GMV includes VAT, excludes returns and discounts offered to customers for the reporting period. GMV is not Revenue of the Company; (10) IAS 17 EBITDA margin in 2017-2019;  
Source: Company data, GfK "Consumer Electronics industry report" (Jan 2021), M&P ANALYTICS LLC, Millward Brawn A/R M/I-Marketing



**Взаимодействие  
с клиентами**

**Машинное обучение**

**Алгоритмы  
умного поиска**

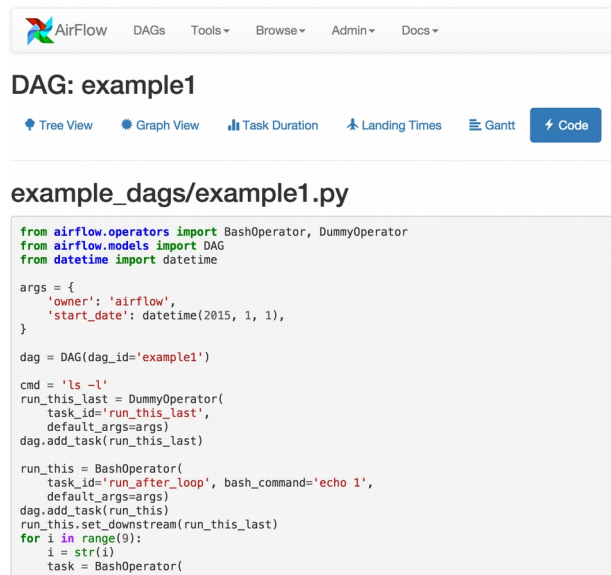
**Pick Up using chat bot  
and machine learning**

**Цифровизация  
бизнес-процессов**

# Проблемы продуктивизации

- Нетипичное описание пайплайнов для DevOps-инженеров
- Организация хранения и версионирования артефактов
- Необходимость типизации потоков данных

# Нетипичное описание пайплайнов для DevOps-инженеров



The screenshot shows the Apache Airflow web interface. At the top, there's a navigation bar with links for DAGs, Tools, Browse, Admin, and Docs. Below this, the title "DAG: example1" is displayed. A row of icons allows switching between different views: Tree View, Graph View, Task Duration, Landing Times, Gantt, and a Code button. The Code button is active, showing the Python code for the DAG. The code defines a DAG named 'example1' with a start date of 2015-01-01. It includes a task 'run\_this\_last' of type DummyOperator and a task 'run\_this' of type BashOperator. The 'run\_this' task is configured to run a bash command 'echo 1' and is scheduled to run 9 times in a loop.

```
from airflow.operators import BashOperator, DummyOperator
from airflow.models import DAG
from datetime import datetime

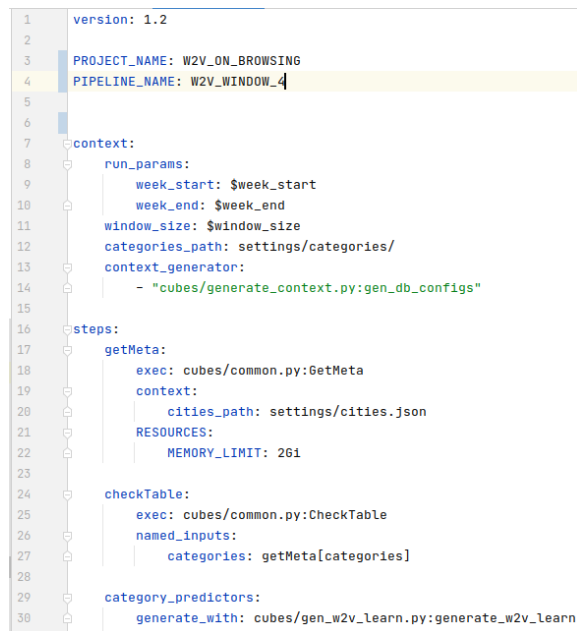
args = {
    'owner': 'airflow',
    'start_date': datetime(2015, 1, 1),
}

dag = DAG(dag_id='example1')

cmd = 'ls -l'
run_this_last = DummyOperator(
    task_id='run_this_last',
    default_args=args)
dag.add_task(run_this_last)

run_this = BashOperator(
    task_id='run_after_loop', bash_command='echo 1',
    default_args=args)
dag.add_task(run_this)
run_this.set_downstream(run_this_last)
for i in range(9):
    i = str(i)
    task = BashOperator(
```

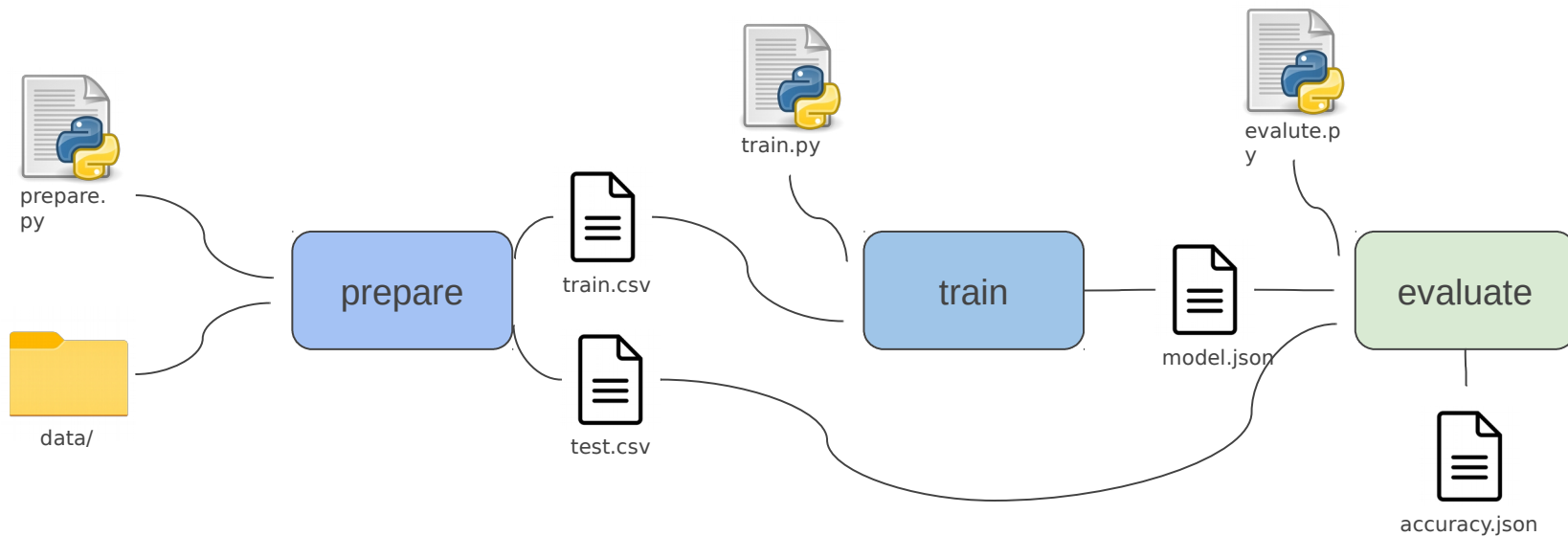
VS



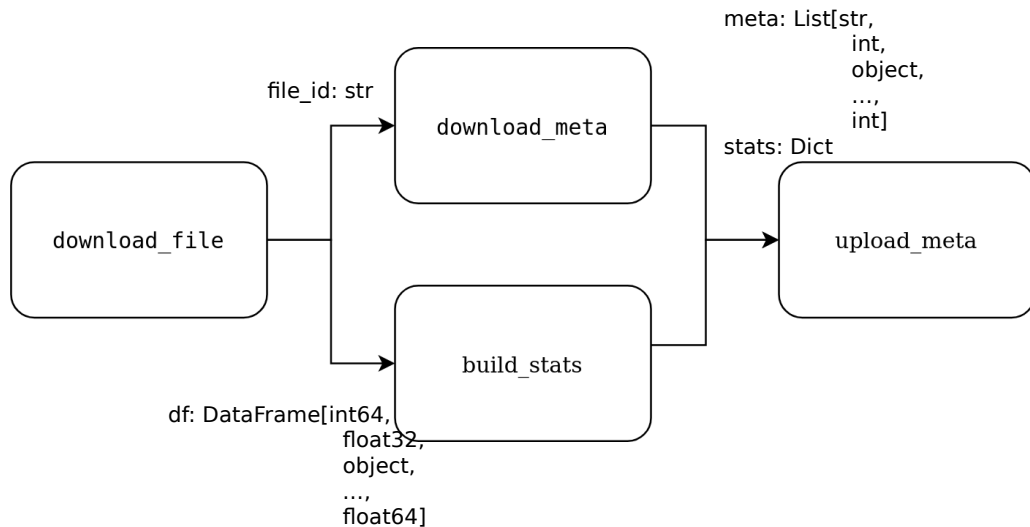
The screenshot shows a YAML configuration for a pipeline. It starts with a version number 1.2. The PROJECT\_NAME is 'W2V\_ON\_BROWSING' and the PIPELINE\_NAME is 'W2V\_WINDOW\_4'. The configuration includes a 'context' section with 'run\_params' (week\_start, week\_end, window\_size, categories\_path) and a 'context\_generator' pointing to 'cubes/generate\_context.py:gen\_db\_configs'. The 'steps' section includes 'getMeta' (executing 'cubes/common.py:GetMeta' with context and resources like 'cities\_path' and 'MEMORY\_LIMIT'), 'checkTable' (executing 'cubes/common.py:CheckTable' with named inputs), and 'category\_predictors' (executing 'cubes/gen\_w2v\_learn.py:generate\_w2v\_learn').

```
1 version: 1.2
2
3 PROJECT_NAME: W2V_ON_BROWSING
4 PIPELINE_NAME: W2V_WINDOW_4
5
6
7 context:
8   run_params:
9     week_start: $week_start
10    week_end: $week_end
11    window_size: $window_size
12    categories_path: settings/categories/
13    context_generator:
14      - "cubes/generate_context.py:gen_db_configs"
15
16 steps:
17   getMeta:
18     exec: cubes/common.py:GetMeta
19     context:
20       cities_path: settings/cities.json
21     RESOURCES:
22       MEMORY_LIMIT: 261
23
24   checkTable:
25     exec: cubes/common.py:CheckTable
26     named_inputs:
27       categories: getMeta[categories]
28
29   category_predictors:
30     generate_with: cubes/gen_w2v_learn.py:generate_w2v_learn
```

# Хранение и версионирование артефактов



# Типизация потоков данных



# В итоге

Реализовано и активно используется в production решение, которое покрывает все перечисленные проблемы:

- Поддержку ML-пайплайнов осуществляют DevOps-инженеры
- Хранение и версионирование артефактов реализуются незаметно для ML-разработчиков
- В продакшене используются типизация потоков данных, в том числе DataFrame`ов (для дева типизация опциональна)



# Спасибо за внимание!

По возникшим вопросам можно обратиться по адресам:

[Yuliya.Lavrushkina@mvideo.ru](mailto:Yuliya.Lavrushkina@mvideo.ru)

[Evgeny.Vorobey@mvideo.ru](mailto:Evgeny.Vorobey@mvideo.ru)