

АЛЬПАКА И КАПИБАРЫ

Лев Глейзер

#MEDIASCOPECONF

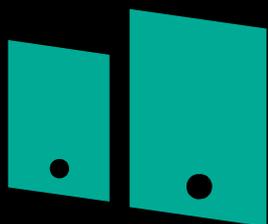
DATA. WE TAKE IT PERSONALLY



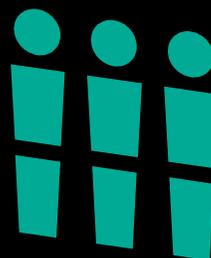
СТРАТЕГИЯ

НОВАЯ ИДЕОЛОГИЯ

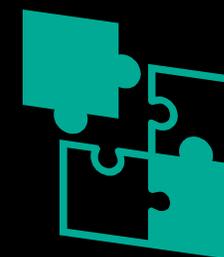
MOBILE FIRST



SYNTHETIC



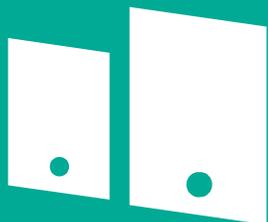
DATAFLOW



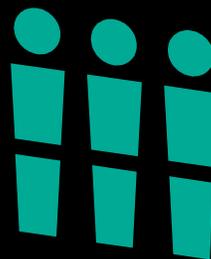
СТРАТЕГИЯ

НОВАЯ ИДЕОЛОГИЯ

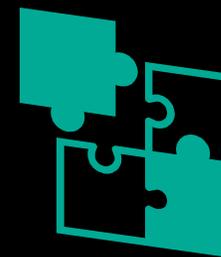
MOBILE FIRST



SYNTHETIC



DATAFLOW

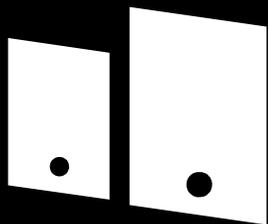


#MEDIASCOPECONF

СТРАТЕГИЯ

НОВАЯ ИДЕОЛОГИЯ

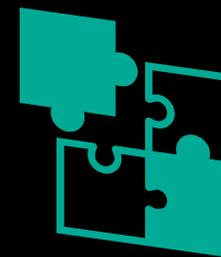
MOBILE FIRST



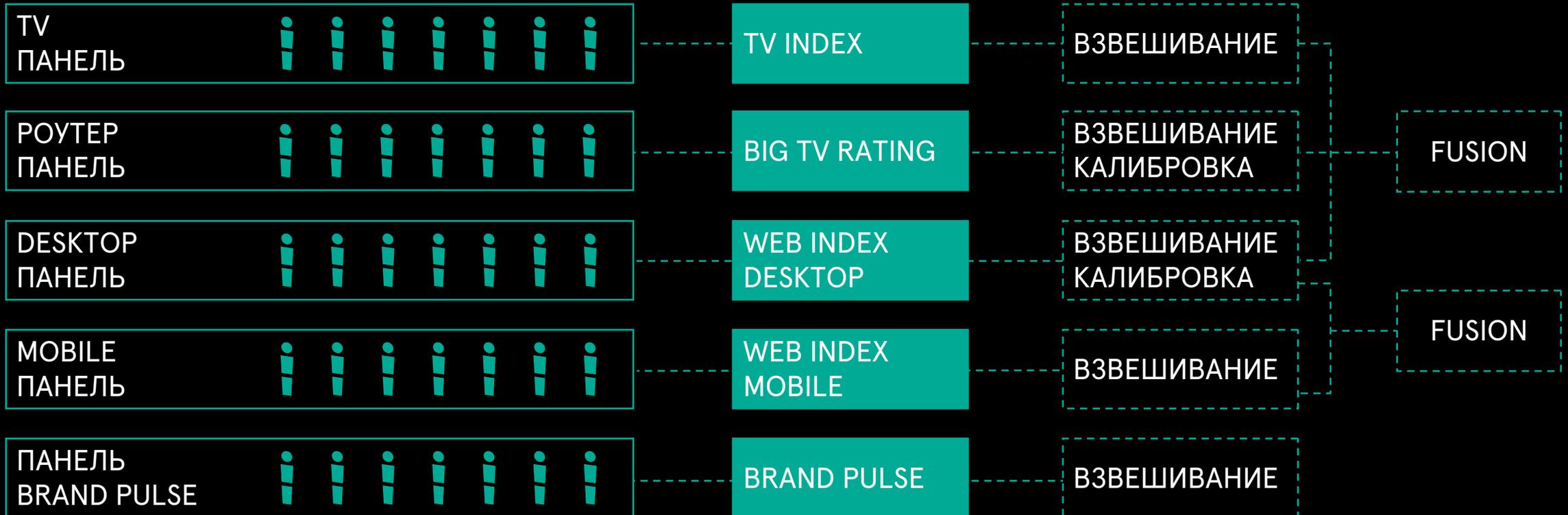
SYNTHETIC



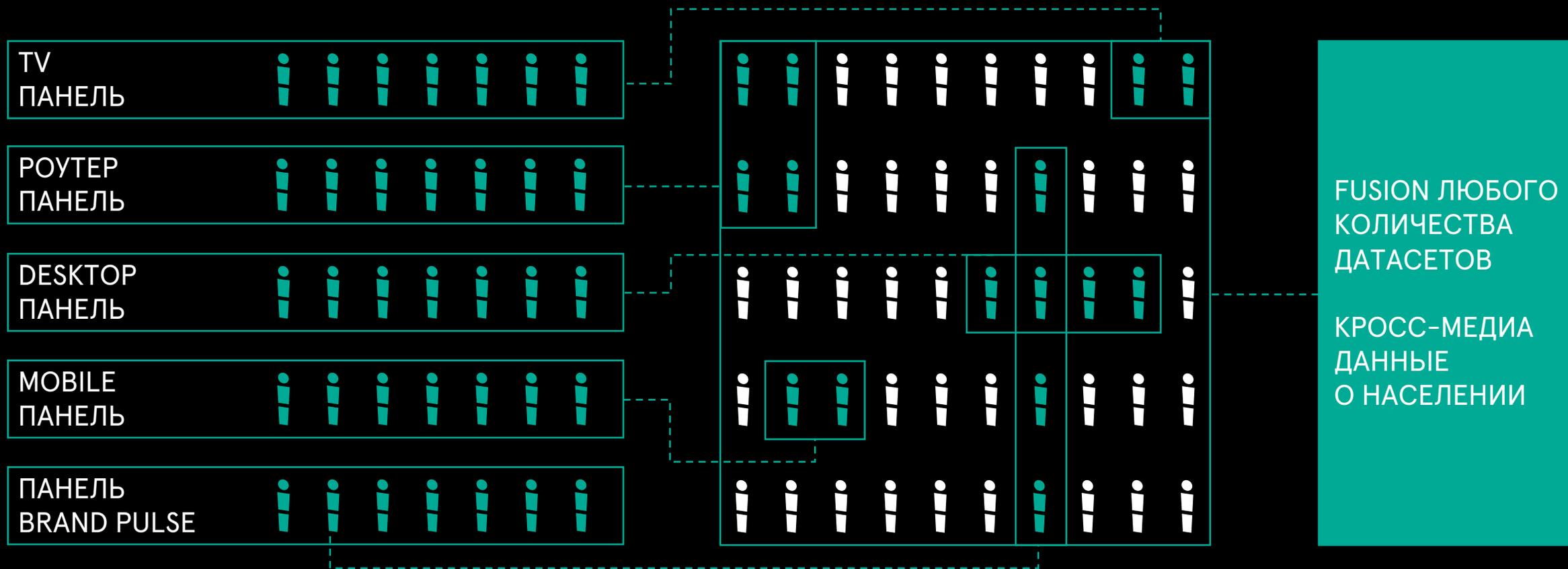
DATAFLOW



КАК ЭТО УСТРОЕНО СЕЙЧАС

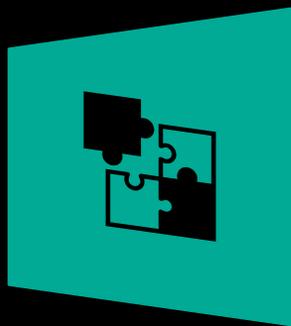


СИНТЕТИЧЕСКАЯ ПАНЕЛЬ



СИНТЕТИКА

CROSS INDEX



ПЕРВЫЙ ПРОЕКТ, ОСНОВАННЫЙ НА СИНТЕТИКЕ

В ОСНОВЕ – СОЦИОЛОГИЯ (ПАНЕЛЬ + CENSUS)

MOBILE FIRST

ОБЪЕДИНЕНИЕ МЕДИА

01/

ПЕРВАЯ РАЗРАБОТКА:

100К ПАНЕЛИСТОВ НА ОСНОВЕ ДАННЫХ ПО ТВ И ИНТЕРНЕТУ

02/

ОБЪЕДИНЕНИЕ ДАННЫХ С УЧЕТОМ ОПЫТА FUSION И КАЛИБРОВОК

03/

ЧАСТИЧНЫЕ ПЕРЕСЕЧЕНИЯ МЕЖДУ СУЩЕСТВУЮЩИМИ ПАНЕЛЯМИ

04/

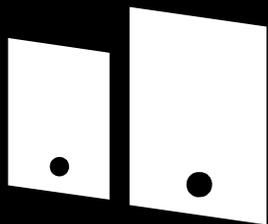
ЗАГРУЗКА ДАННЫХ В ДАТА-ПЛАТФОРМУ

#MEDIASCOPECONF

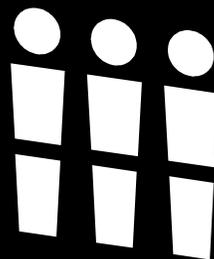
СТРАТЕГИЯ

НОВАЯ ИДЕОЛОГИЯ

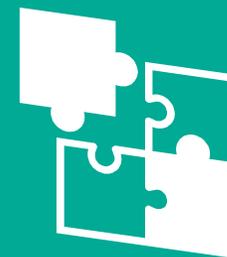
MOBILE FIRST



SYNTHETIC



DATAFLOW



КОНЦЕПТ



EVENT LOG

ЖУРНАЛ СОБЫТИЙ

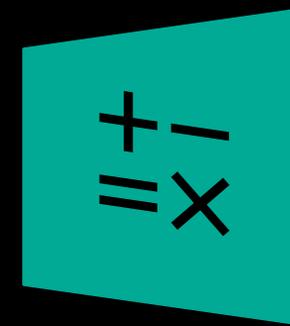
ПОЛНАЯ ИНФОРМАЦИЯ О КОНТАКТЕ:
КТО, ЧТО, ГДЕ, КОГДА



DATA SET

ПОСТАВКА БОЛЬШОГО
ЧИСЛА РАЗНЫХ ДАТАСЕТОВ

ПРЕДОСТАВЛЕНИЕ ШИРОКОГО СПЕКТРА
ВОЗМОЖНОСТЕЙ ДЛЯ ИХ АНАЛИЗА

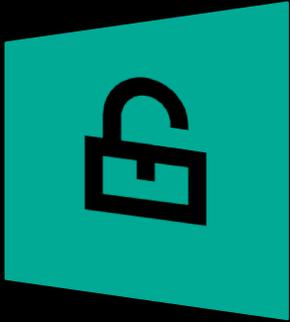


МАТЕМАТИКА КАК СЕРВИС

УНИВЕРСАЛЬНЫЕ
СЕРТИФИЦИРОВАННЫЕ
ПРОЦЕДУРЫ

FAAS (FUSION AS A SERVICE),
ВЗВЕШИВАНИЕ, КАЛИБРОВКА, ETC.

ПРЕИМУЩЕСТВА



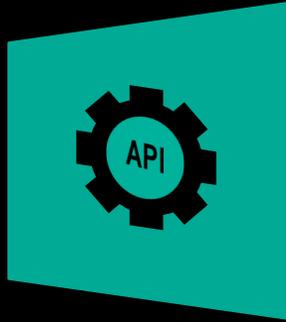
OPEN SOURCE

HORTONWORKS COMMUNITY EDITION
+ APACHE SPARK + PYTHON

СОВМЕСТИМОСТЬ

НЕЗАВИСИМОСТЬ

РАСПРОСТРАНЕННОСТЬ



OPEN API

ОТКРЫТАЯ МОДЕЛЬ ДАННЫХ
ОПУБЛИКОВАННОЕ ОПИСАНИЕ
АЛГОРИТМОВ РАСЧЕТОВ И
ПРЕОБРАЗОВАНИЙ

ОТКРЫТЫЙ ПРОГРАММНЫЙ ИНТЕРФЕЙС
МАТЕМАТИЧЕСКИХ СЕРВИСОВ



CLOUD SERVICES

ЭЛАСТИЧНОЕ РАСШИРЕНИЕ
РЕСУРСОВ ON-DEMAND

ОБЛАКО HUAWEI НА ТЕРРИТОРИИ
РОССИИ (ОПЕРАТОР SBERCLOUD)

ДОСТУП К ДАННЫМ И СЕРВИСАМ
ПЛАТФОРМЫ В РЕЖИМЕ
PRIVATE CLOUD

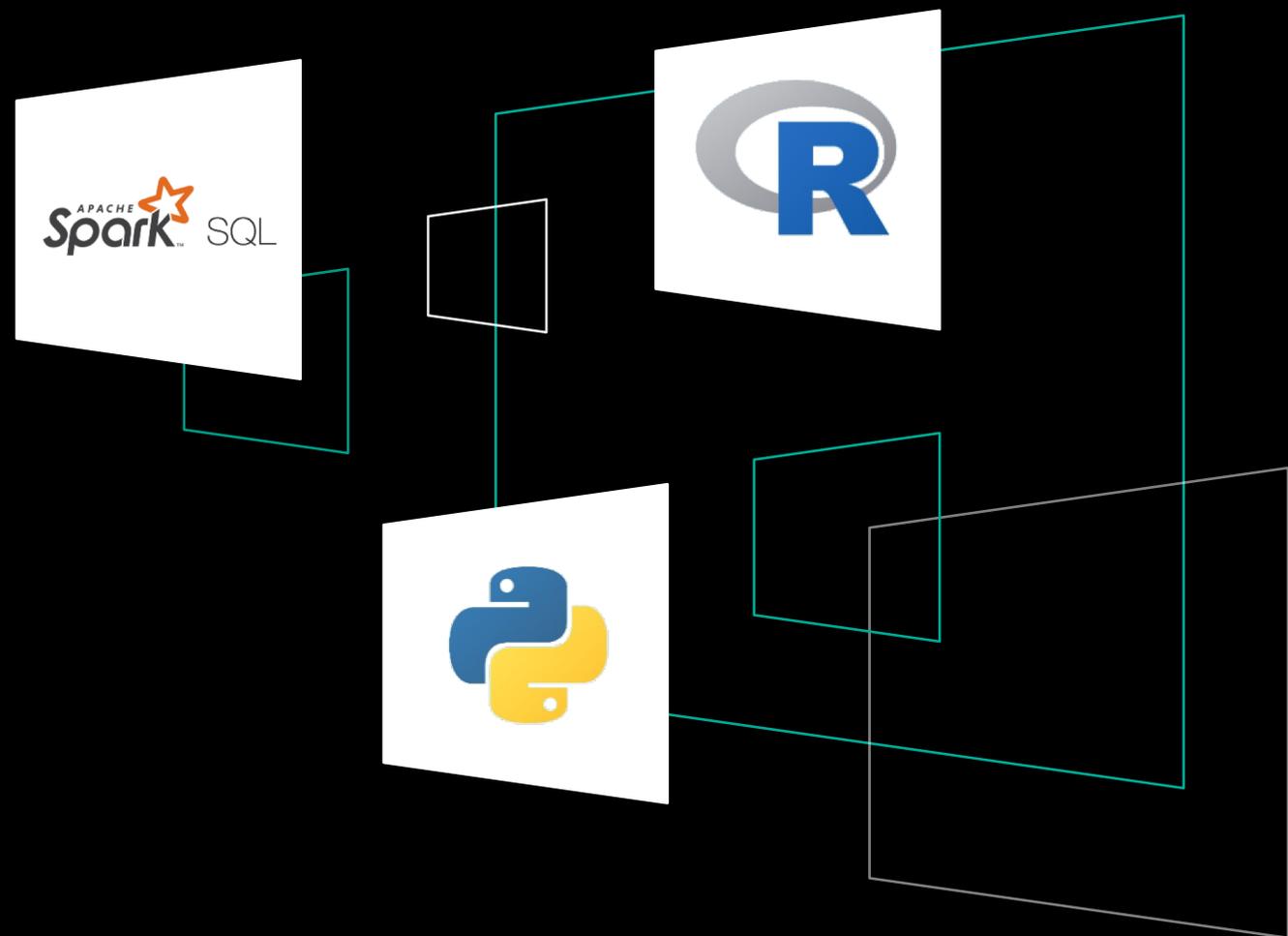
НОВЫЙ ФОРМАТ ПОСТАВКИ

01/

БОЛЬШЕ ДАННЫХ – ВЫШЕ ТРЕБОВАНИЯ
К СКОРОСТИ ОБРАБОТКИ

02/

ИСПОЛЬЗОВАНИЕ ДАННЫХ ПЕРЕХОДИТ
В ПРОФЕССИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ ЯЗЫК



НОВЫЙ ФОРМАТ ПОСТАВКИ



ВЧЕРА

ПОЛЬЗОВАТЕЛЬСКОЕ ПО

ОТЧЕТЫ

ДЭШБОРДЫ

СЕГОДНЯ

ПОЛЬЗОВАТЕЛЬСКОЕ ПО

ДЭШБОРДЫ

ОТЧЕТЫ

API

ЗАВТРА

API

DATALAB

ПОЛЬЗОВАТЕЛЬСКОЕ ПО

ДЭШБОРДЫ

ОТЧЕТЫ

DATALAB

Audience

Задача посчитать аудиторию проектов:

- Avito.ru – аудитория одного проекта
- Avito.ru, lvi.ru – аудиторию 2-х проектов, т.е. количество людей, которые посещают оба сайта
- Total – Total Internet

```

...
facility = 'mobile' # возможные значения: 'desktop', 'mobile', 'desktop-mobile'
rcats = rc.ResponsumCats(facility)
rtask = rt.ResponsumTask(facility)
# задаем период расчета
date_from = '2020-05-01'
date_to = '2020-05-31'
# задаем Типы пользования Интернетом
usetypes = rcats.get_usetype('all')
# задаем Население, нас интересует города с населением 100тыс. и больше
population = rcats.get_population('Russia100+')
# задаем Возрастные группы, нас интересует 12+
ages = rcats.get_age_groups('12+')

```

Расчет аудитории по проекту Avito.ru

```

# Задаем название проекта для отображения в DataFrame
project_name = 'avito.ru'
# Задаем медиа фильтр, в нашем случае это ID проекта Avito.ru
media_filter = "site = 15828"
# Задаем список статистик для расчета
statistics=["UnwReach", "Reach", "ADR", "AWR", "Uni", "ReachPer", "StatReach"]
# Указываем порядок группировки, в нашем случае ее нет

```

jupyter duplication (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Графики

```

In [ ]:
1 import plotly.express as px
2 import numpy as np
3 img_rgb = np.array([[[255, 0, 0], [0, 255, 0], [0, 0, 255]],
4                    [[0, 255, 0], [0, 0, 255], [255, 0, 0]]],
5                    dtype=np.uint8)
6 # fig = px.imshow(df_dup)
7 fig = px.density_heatmap(df_dup, x="media_site", y="duplication_site", z="stat_reach")
8 fig.show()

```

```

In [34]:
1 df_dup.plot(kind='bar', stacked=True)

```

Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x7f80a2517dd0>

```

(rtask.get_result(task_audience), project_name=project_name)

```

38f22028d7d2> поступила в обработку
=====] время расчета: 0:00:55.706044

reach	unreach	adr_per	awr_per	awr	adr
46592.158751	10385.0	8.157812	21.737146	26600.286646	9982.918132

История TotalInternet

Таблицы

Вывести результаты в один DataFrame

```

df_avito_ivi, df_total])

```

reach	unreach	adr_per	awr_per	awr	adr
39969.333508	9007.0	7.307620	19.022255	23278.007107	8942.516823
46592.158751	10385.0	8.157812	21.737146	26600.286646	9982.918132
95336.209469	19812.0	69.766854	76.068995	93087.523456	85375.436012

```

df.get_excel_filename('audience-avito-ivi')
df.info()
df.to_excel('report', index=False)
df.to_excel('report', index=False)

```

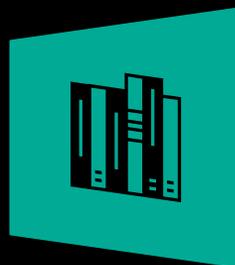
ЧТО ПОЛУЧАТ ПОЛЬЗОВАТЕЛИ

01.



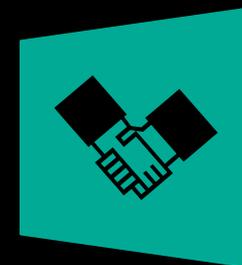
ТЕХНИЧЕСКИЕ МОЩНОСТИ –
ВЫДЕЛЕННЫЕ ВИРТУАЛКИ

02.



ПОЛНЫЙ НАБОР ОПИСАТЕЛЬНОЙ
ДОКУМЕНТАЦИИ И ПРОСТРАНСТВА
В CONFLUENCE

03.



СОПРОВОЖДЕНИЕ
ЭКСПЕРИМЕНТАЛЬНЫХ
ПРОЕКТОВ

ЧТО ЕСТЬ УЖЕ СЕЙЧАС

ДАТАСЕТЫ

ТЕЛЕВИЗИОННЫЕ ДАННЫЕ

ДААННЫЕ SITE-CENTRIC ПАНЕЛИ

ДААННЫЕ USER-CENTRIC ПАНЕЛИ

MOBILE

BRAND PULSE

ТЕХНОЛОГИЧЕСКИЙ СТЭК



СОВМЕСТНАЯ РАБОТА

ПОЛЬЗОВАТЕЛИ

НАЦИОНАЛЬНЫЙ
РЕКЛАМНЫЙ АЛЬЯНС

АГЕНТСКИЕ ГРУППЫ

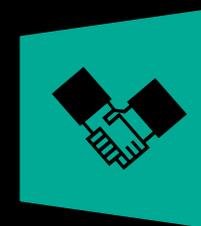
ПОДДЕРЖКА MEDIASCOPE



ТЕХНИЧЕСКАЯ
ПОМОЩЬ И ОБУЧЕНИЕ



ВЫРАБОТКА
СТАНДАРТОВ РЫНКА



СОВМЕСТНОЕ
РАЗВИТИЕ

СПАСИБО!

Лев Глейзер,
директор по информационным технологиям

