Выгрузка расходов из Яндекс Директ в BQ

- Алгоритм создания потока из системы Я.Директ в BQ:
  - 1) Источник данных > Приемник данных
  - 2) Настройка источника данных
  - 3) Настройка приемника данных
  - 4) Общие настройки

На данный момент доступно создание потока по выгрузке расходов из системы Яндекс Директ в Google BigQuery.

Из-за особенностей API Яндекс отдает расходы двумя отчетами:

- Статистика. В этом отчете передается вся информация о расходах
- Контент. В этом отчете передаются url объявления

Для получения и объединения двух этих отчетов необходимо использовать определенный набор данных, подробнее в первом шаге настроек потока.

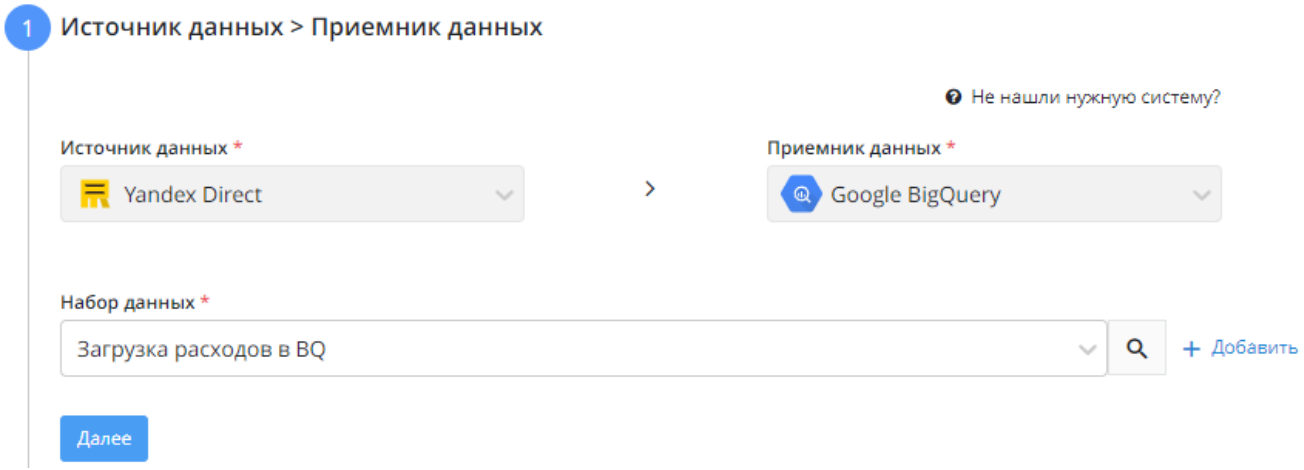
Первоначально необходимо зайти в систему Garpun Feeds https://feeds.garpun.com/ под своим логином и паролем и нажать кнопку "+Поток данных".

Алгоритм создания потока из системы Я.Директ в BQ:

1) Источник данных > Приемник данных

В открывшемся меню выбираем:

- Источник данных - Yandex Direct,
- Приемник данных - Google BQ
- Набор данных "Загрузка расходов в BQ". Преимущество этого набора данных в том, что в результирующую таблицу будут добавлены поля utm_cource, utm_medium, utm_campaign, utm_term и utm_content, содержащие соответствующие метки. Это поможет более детально анализировать трафик и расходы.



2) Настройка источника данных

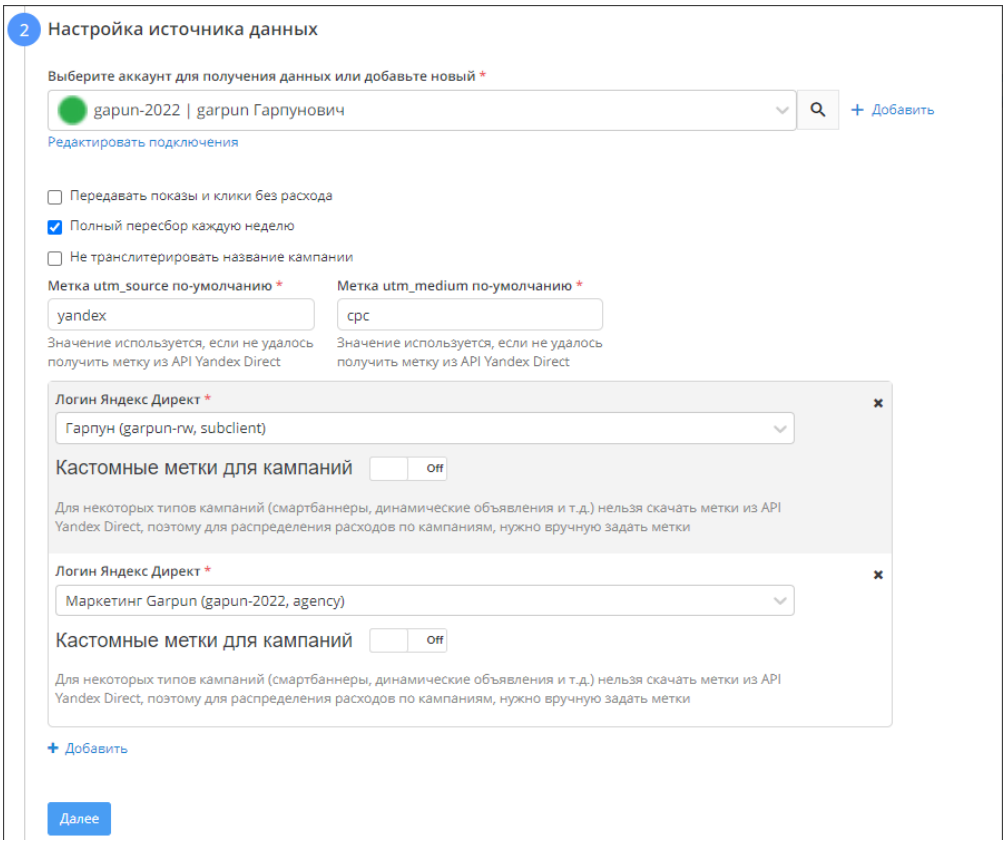
- Выбираем необходимое подключение, либо добавляем новое
- Указываем логин рекламного аккаунта, с которого собирается статистика. **По нажатию на кнопку "+ Добавить" можно добавлять логины, однако просим учесть, что каждый новый логин будет тарифицироваться как отдельный поток, т.к. для каждого логина нужно отправлять отдельный запрос в Яндекс.**

Необязательные настройки:

Необходим если нужно выкачивать всю статистику. По умолчанию система игнорирует ситуации когда по объявлениям нет расходов, даже если есть показы и клики. Но в некоторых случаях это может приводить к потере данных, именно для таких случаев была сделана данная настройка.

Настройка устанавливает полный пересбор раз в неделю. Полезно для кампаний где происходят частые изменения. Так же полезно в случаях использования объемных кампаний с большими расходами. Яндекс обладает своей внутренней системой защиты. Алгоритм проверяет различные условия и убирает из расходов ботов и "скликивание". Это происходит в течении 30 дней после формирования отчета в самом Яндексе, поэтому рекомендуется проверять статистику каждый месяц, так как значения в Яндексе спустя 30 дней могут оказаться меньше, чем значения полученные в первый день
В API Яндекс.Директ не были реализованы методы получения контента объявлений для кампаний, созданных через мастер-кампаний. Для того, чтоб избежать not-set в статистике, берется название кампании из статистики и к этому названию применяется алгоритм транслитерации.

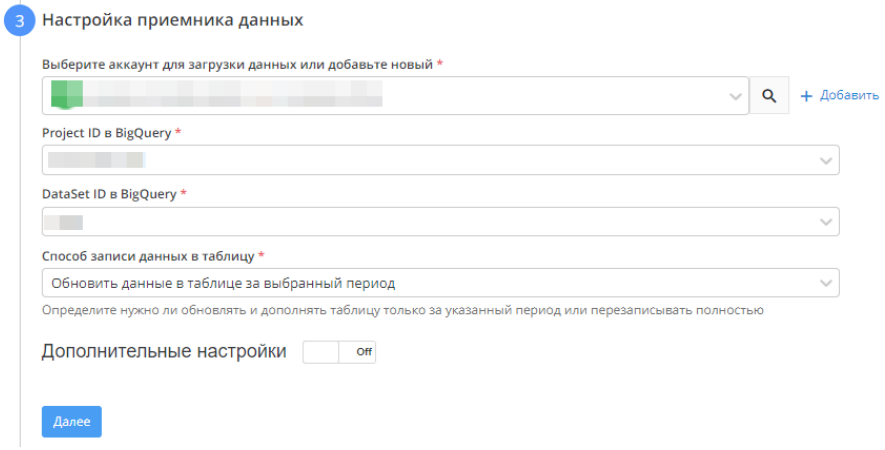
Алгоритм транслитерации запускается по умолчанию, вне зависимости латиницией или кириллицей записано название кампании. Для того, чтоб избежать "некрасивых" названий, реализована опция отключения транслитерации, но обязательно нужно убедиться, что все названия не требуется переводить в формат, который принимает Аналитика, т.е. названия кампаний на латинице.
Метки так же используются, если по каким-либо причинам мы не получаем эти метки из API Директа. Их необходимо проставить, чтобы система могла использовать их когда нужно привязать расходы без меток. Если говорить кратко, то все расходы, по которым мы не получили метки из Директа будут промечены указанными значениями. По умолчанию используется utm_source=yandex, utm_medium=cpc
Для некоторых типов кампаний (смартбаннеры, динамические объявления и т.д.) нельзя скачать метки из API Yandex Direct, поэтому, для распределения расходов по кампаниям, нужно вручную задать метки. Это так же касается объявлений мастера кампаний.
По всем этим кампаниям мы не получаем контент объявлений. Без контента, не имея ссылок из объявления, система не может разбить данные по меткам. Чтобы избежать потери данных, алгоритм берет название кампании из Яндекса, транслитерирует его, использует полученное имя как метку и передает данные в Аналитику.
Чтобы избежать транслитерации нужно использовать кастомные метки, либо отключить алгоритм транслитерации при помощи соответствующей опции.



3) Настройка приемника данных

- Выбираем подключение Google BQ, либо добавляем подключение через кнопку "+Добавить"
- Указываем Project ID в BigQuery
- Указываем dataset. Не обязательно создавать Dataset вручную в самом BigQuery, можно ввести название вручную - Dataset будет создан автоматически при запуске потока.

- В качестве способа записи данных в таблицу оставляем обновление



4) Общие настройки

- В графе “Название потока” ввести название либо оставить сгенерированное автоматически
- В графе “Период сбора при автоматическом запуске” можно выбрать за какой период поток будет пересобирать статистику. По умолчанию - 30 дней.
- В графе “Расписание” - выбрать например 7:00 утра, в это время поток будет запускаться ежедневно
- Нажать "Готово"

В качестве итоговой таблицы мы получаем таблицу со следующей схемой:

Название	Тип
ad_format	String
ad_group_id	Int64
ad_group_name	String
ad_id	Int64
ad_network_type	String
age	String
avg_click_position	Int64
avg_effective_bid	Int64
avg_impression_position	Float32
campaign_id	Int64
campaign_name	String
campaign_url_path	String
campaign_type	String
clicks	Int64
cost	Float64
criteria	String
criteria_id	Int64
criteria_type	String

date	Date
device	String
gender	String
impressions	Int64
location_of_presence_id	Int64
location_of_presence_name	String
match_type	String
mobile_platform	String
placement	String
rl_adjustment_id	Int64
slot	String
targeting_category	String
targeting_location_id	Int64
targeting_location_name	String
yandex_client_login	String
source	String
medium	String
campaign	String
term	String
content	String