

Загрузка расходов из Yandex Direct в ClickHouse

- Алгоритм создания потока из системы Я.Директ в ClickHouse:
 - 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 ClickHouse.

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".

Обратите внимание

Для корректной передачи расходов требуется версия ClickHouse не ниже 23.8. К сожалению, версии ниже 23.8 не поддерживают часть функций, в следствие чего агрегирующие запросы могут работать некорректно.

Алгоритм создания потока из системы Я.Директ в ClickHouse:

1) Источник данных > Приемник данных

В открывшемся меню выбираем:

- Источник данных - Yandex Direct,
- Приемник данных - Database (PostgreSQL, MySQL, ClickHouse)
- Набор данных "Загрузка расходов СН". Преимущество этого набора данных в том, что в результирующую таблицу будут добавлены поля utm_source, utm_medium, utm_campaign, utm_term и utm_content, содержащие соответствующие метки. Это поможет более детально анализировать трафик и расходы.

1 Источник данных > Приемник данных

Не нашли нужную систему?

Источник данных *
Yandex Direct

Приемник данных *
Database (PostgreSQL, MySQL, ClickHouse)

Набор данных *
Загрузка расходов СН

Далее

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:

Необходимо 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 привести 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 pot-set в статистике, берется название кампании из статистики и к этому названию применяется алгоритм транслитерации.

Алгоритм транслитерации запускается по умолчанию, вне зависимости латиницей или кириллицей записано название кампании. Для того, чтоб избежать "некрасивых" названий, реализована опция отключения транслитерации, но обязательно нужно убедиться, что все названия не требуются переводить в формат, который требуется для корректной передачи, 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.

2 Настройка источника данных

Выберите аккаунт для получения данных или добавьте новый *

garpun-2022 | garpun Гарпунович + Добавить

[Редактировать подключения](#)

Передавать показы и клики без расхода

Полный пересбор каждую неделю

Не транслитерировать название кампании

Метка `utm_source` по-умолчанию * Значение используется, если не удалось получить метку из API Yandex Direct

Метка `utm_medium` по-умолчанию * Значение используется, если не удалось получить метку из API Yandex Direct

Логин Яндекс Директ *

Кастомные метки для кампаний off

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

Логин Яндекс Директ *

Кастомные метки для кампаний off

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

[+ Добавить](#)

[Далее](#)

3) Настройка приемника данных

- Выбираем подключение ClickHouse, либо добавляем подключение через кнопку "+Добавить"

Для создания нового подключения ClickHouse необходимо нажать кнопку "+Добавить" на третьем шаге настроек потока, либо на странице подключений(<https://account.garpun.com/page?p=5184&a=2>) выбрать "Database (PostgreSQL, MySQL, ClickHouse)" и нажать

+ ПОДКЛЮЧЕНИЕ

кнопку

Дальнейшие действия идентичны для обоих вариантов:

- Вводим Имя хоста, или IP-адрес
 - Порт(опционально)
 - Имя БД
 - Имя пользователя(опционально)
 - Пароль(опционально)
 - Тип БД - выбираем "ClickHouse"
 - Использовать SSL подключение? – выбираем "Да"
- Выбираем базу данных
 - Название схемы БД заполнять не нужно, т.к. это важно только для передачи в PostgreSQL.
 - В качестве способа записи данных в таблицу оставляем обновление

3 Настройка приемника данных

Выберите аккаунт для загрузки данных или добавьте новый *

Название базы данных *

ВАЖНО. Для корректной работы с Clickhouse требуется версия >= 22.5

Название схемы базы данных

Поле необходимо заполнить, если Ваша база данных PostgreSQL

Способ записи данных в таблицу *

Определите нужно ли обновлять и дополнять таблицу только за указанный период или перезаписывать полностью - опция 'Обновить данные в таблице за выбранный период' требует заполненного поля партиционирования при создании таблицы

Дополнительные настройки для Clickhouse

On

Создавать реплицированные таблицы для Clickhouse

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:

date
client_login
clicks
impressions
cost

campaign_id
campaign_name
campaign_name_lat
ad_id
ad_group_name
keyword
device_type
phrase_id
position_type
position
gbid
source
source_type
addphrases
retargeting_id
CriteriaType
CampaignType
MatchType
Slot
AdNetworkType
Age
CarrierType
ExternalNetworkName
Gender
MobilePlatform
AvgClickPosition
AvgCpc
AvgImpressionPosition
AvgPageviews
BounceRate
Bounces
ConversionRate
Conversions
CostPerConversion
Ctr
GoalsRoi
Revenue
Sessions

RIAdjustmentId
CriterionId
AdGroupId
Placement
LocationOfPresenceName
LocationOfPresenceId
Criterion
url
utm_content
utm_campaign
utm_medium
utm_term
utm_source



Обратите внимание

Из-за технических особенностей передачи данных к итоговой таблице могут быть добавлены префиксы систем, участвующих в выгрузке, указание набора данных, а так же название кабинета/аккаунта. Так, при выгрузке расходов из системы Яндекс.Директ в ClickHouse нейминг таблицы будет следующим: db_yandex_direct_keyword_cost_upload_ch_our_cabinet, где:

db – система-приемник;

yandex_direct – система-источник;

keyword_cost_upload_ch – название набора данных;

our_cabinet – название кабинета Яндекс.Директ